



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 102649

To: Jennifer Graser

Location: CM1-7E12

Art Unit: 1645

Friday, September 05, 2003

Case Serial Number: 10/049473

From: Beverly Shears

Location: Biotech-Chem Library

CM1-1E05

Phone: 308-4994

beverly.shears@uspto.gov

Search Notes

8/99
Rehoot
Hermanns

98/8930
wo/8/8930
P.55
P114-116
ref. 5021
Interference

BEST AVAILABLE COPY

THIS PAGE BLANK (USPTO)

SEARCH REQUEST FORM

Requestor's

Name: _____

Serial

Number: _____

Date: _____

Phone: _____

Art Unit: _____

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

STAFF USE ONLY

Date completed: 09-05-03

Searcher: Beverly 29994

Terminal time: 25

Elapsed time: _____

CPU time: _____

Total time: 25

Number of Searches: _____

Number of Databases: 1

Search Site

_____ STIC

_____ CM-1

_____ Pre-S

Type of Search

_____ N.A. Sequence

_____ A.A. Sequence

_____ Structure

_____ Bibliographic

Vendors

_____ IG

_____ STN

_____ Dialog

_____ APS

_____ Geninfo

_____ SDC

_____ DARC/Questel

☒ Other CGN

BEST AVAILABLE COPY

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2003, 10:33:59 ; Search time 378 Seconds

(Without alignments)
741.416 Million cell updates/sec

Title: US-10-049-473a-2

Perfect score: 1588
Sequence: 1 SNTYKGVDMKKLAGAIR.....IFQYIGGDSSTSSSTSSNE 322

Scoring table:

BLOSUM62
Gapop 10.0 , Gapect 0.5

Searched: 5580241 segs, 870357830 residues

Total number of hits satisfying chosen parameters: 5580241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Pending_Patents_AA.Main:*

1: /cgn2_6/ptodata/1/paa/PCFUS.COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06.COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US07.COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US08.COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US081.COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US082.COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US083.COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US084.COMB.pep.*
9: /cgn2_6/ptodata/1/paa/US085.COMB.pep.*
10: /cgn2_6/ptodata/1/paa/US086.COMB.pep.*
11: /cgn2_6/ptodata/1/paa/US087.COMB.pep.*
12: /cgn2_6/ptodata/1/paa/US088.COMB.pep.*
13: /cgn2_6/ptodata/1/paa/US089.COMB.pep.*
14: /cgn2_6/ptodata/1/paa/US090.COMB.pep.*
15: /cgn2_6/ptodata/1/paa/US091.COMB.pep.*
16: /cgn2_6/ptodata/1/paa/US092.COMB.pep.*
17: /cgn2_6/ptodata/1/paa/US093.COMB.pep.*
18: /cgn2_6/ptodata/1/paa/US094.COMB.pep.*
19: /cgn2_6/ptodata/1/paa/US095.COMB.pep.*
20: /cgn2_6/ptodata/1/paa/US096.COMB.pep.*
21: /cgn2_6/ptodata/1/paa/US097A.COMB.pep.*
22: /cgn2_6/ptodata/1/paa/US097B.COMB.pep.*
23: /cgn2_6/ptodata/1/paa/US098.COMB.pep.*
24: /cgn2_6/ptodata/1/paa/US099A.COMB.pep.*
25: /cgn2_6/ptodata/1/paa/US099B.COMB.pep.*
26: /cgn2_6/ptodata/1/paa/US100.COMB.pep.*
27: /cgn2_6/ptodata/1/paa/US101.COMB.pep.*
28: /cgn2_6/ptodata/1/paa/US102.COMB.pep.*
29: /cgn2_6/ptodata/1/paa/US103.COMB.pep.*
30: /cgn2_6/ptodata/1/paa/US104.COMB.pep.*
31: /cgn2_6/ptodata/1/paa/US60.COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1588	100.0	322	26	US-10-049-473a-2
2	1541	97.0	313	22	US-09-769-787-163

Interference ARG

3	1538	96.9	313	19	US-09-583-110-5200	Sequence 5200, Ap
4	1428	89.9	295	15 <th>US-09-107-433-4127</th> <th>Sequence 4127, Ap</th>	US-09-107-433-4127	Sequence 4127, Ap
5	916	57.7	213	13 <th>US-08-961-083-34</th> <th>Sequence 34, Appl</th>	US-08-961-083-34	Sequence 34, Appl
6	916	57.7	213	22 <th>US-09-765-271-34</th> <th>Sequence 34, Appl</th>	US-09-765-271-34	Sequence 34, Appl
7	916	57.7	213	22 <th>US-09-765-272-34</th> <th>Sequence 34, Appl</th>	US-09-765-272-34	Sequence 34, Appl
8	916	57.7	213	22 <th>US-09-765-272-34</th> <th>Sequence 34, Appl</th>	US-09-765-272-34	Sequence 34, Appl
9	634	39.9	344	20 <th>US-10-417-884-6355</th> <th>Sequence 6355, Ap</th>	US-10-417-884-6355	Sequence 6355, Ap
10	608.5	38.3	368	15 <th>US-09-134-000-4284</th> <th>Sequence 4284, Ap</th>	US-09-134-000-4284	Sequence 4284, Ap
11	608.5	38.3	368	15 <th>US-10-434-665-4284</th> <th>Sequence 4284, Ap</th>	US-10-434-665-4284	Sequence 4284, Ap
12	608.5	38.3	368	15 <th>US-10-434-665-4284</th> <th>Sequence 4284, Ap</th>	US-10-434-665-4284	Sequence 4284, Ap
13	605.5	38.1	342	28 <th>US-10-206-576-182</th> <th>Sequence 182, App</th>	US-10-206-576-182	Sequence 182, App
14	588.5	37.1	322	30 <th>US-10-417-884-6113</th> <th>Sequence 6113, Ap</th>	US-10-417-884-6113	Sequence 6113, Ap
15	562.5	35.4	302	28 <th>US-10-206-576-184</th> <th>Sequence 184, App</th>	US-10-206-576-184	Sequence 184, App
16	450.5	28.4	302	30 <th>US-10-417-884-4736</th> <th>Sequence 4736, Ap</th>	US-10-417-884-4736	Sequence 4736, Ap
17	375.5	23.6	299	26 <th>US-10-049-473a-3</th> <th>Sequence 3, Appl</th>	US-10-049-473a-3	Sequence 3, Appl
18	373	23.5	299	26 <th>US-10-049-473a-4</th> <th>Sequence 4, Appl</th>	US-10-049-473a-4	Sequence 4, Appl
19	367	23.1	299	26 <th>US-10-049-473a-5</th> <th>Sequence 5, Appl</th>	US-10-049-473a-5	Sequence 5, Appl
20	357.5	22.5	292	22 <th>US-09-791-537-115536</th> <th>Sequence 115536, Appl</th>	US-09-791-537-115536	Sequence 115536, Appl
21	342.5	21.6	292	13 <th>US-08-902-393-11</th> <th>Sequence 11, Appl</th>	US-08-902-393-11	Sequence 11, Appl
22	322	20.3	74	1 <th>PCT-US97-07950-508</th> <th>Sequence 508, App</th>	PCT-US97-07950-508	Sequence 508, App
23	294.5	18.5	371	1 <th>PCT-US02-36123-5490</th> <th>Sequence 5490, Ap</th>	PCT-US02-36123-5490	Sequence 5490, Ap
24	234.5	14.8	309	1 <th>PCT-US02-36123-5490</th> <th>Sequence 4943, Ap</th>	PCT-US02-36123-5490	Sequence 4943, Ap
25	217.5	13.7	328	12 <th>US-08-827-353-4943</th> <th>Sequence 5956, Ap</th>	US-08-827-353-4943	Sequence 5956, Ap
26	217.5	13.7	328	20 <th>US-09-611-529-5956</th> <th>Sequence 5956, Ap</th>	US-09-611-529-5956	Sequence 5956, Ap
27	217.5	13.7	328	25 <th>US-09-950-084-5956</th> <th>Sequence 5192, Ap</th>	US-09-950-084-5956	Sequence 5192, Ap
28	212.5	13.4	320	11 <th>US-08-781-986A-5192</th> <th>Sequence 5192, Ap</th>	US-08-781-986A-5192	Sequence 5192, Ap
29	212.5	13.4	320	13 <th>US-08-956-171-5192</th> <th>Sequence 5192, Ap</th>	US-08-956-171-5192	Sequence 5192, Ap
30	212.5	13.4	320	13 <th>US-08-956-171B-5192</th> <th>Sequence 5192, Ap</th>	US-08-956-171B-5192	Sequence 5192, Ap
31	212.5	13.4	320	13 <th>US-08-956-171C-5192</th> <th>Sequence 5192, Ap</th>	US-08-956-171C-5192	Sequence 5192, Ap
32	212.5	13.4	320	13 <th>US-08-956-171D-5192</th> <th>Sequence 5192, Ap</th>	US-08-956-171D-5192	Sequence 5192, Ap
33	212.5	13.4	320	13 <th>US-08-956-171E-5192</th> <th>Sequence 5192, Ap</th>	US-08-956-171E-5192	Sequence 5192, Ap
34	210	13.2	330	18 <th>US-09-450-969-5963</th> <th>Sequence 5963, Ap</th>	US-09-450-969-5963	Sequence 5963, Ap
35	210	13.2	330	26 <th>US-10-092-411A-3811</th> <th>Sequence 3811, Ap</th>	US-10-092-411A-3811	Sequence 3811, Ap
36	173	10.9	42	12 <th>US-08-832-030-453</th> <th>Sequence 453, App</th>	US-08-832-030-453	Sequence 453, App
37	169.5	10.7	594	22 <th>US-09-791-537-146251</th> <th>Sequence 146251, App</th>	US-09-791-537-146251	Sequence 146251, App
38	163	10.3	621	23 <th>US-09-897-516-7072</th> <th>Sequence 7072, Ap</th>	US-09-897-516-7072	Sequence 7072, Ap
39	163	10.3	621	31 <th>US-06-215-161-7072</th> <th>Sequence 7072, Ap</th>	US-06-215-161-7072	Sequence 7072, Ap
40	159	10.0	677	16 <th>US-09-252-691C-8092</th> <th>Sequence 8092, Ap</th>	US-09-252-691C-8092	Sequence 8092, Ap
41	159	10.0	677	16 <th>US-09-252-691C-8092</th> <th>Sequence 8092, Ap</th>	US-09-252-691C-8092	Sequence 8092, Ap
42	159	10.0	677	30 <th>US-10-417-886-8092</th> <th>Sequence 8092, Ap</th>	US-10-417-886-8092	Sequence 8092, Ap
43	155.5	9.8	282	22 <th>US-09-791-537-120537</th> <th>Sequence 120537, App</th>	US-09-791-537-120537	Sequence 120537, App
44	155.5	9.8	336	22 <th>US-09-791-537-39861</th> <th>Sequence 39861, A</th>	US-09-791-537-39861	Sequence 39861, A
45	147.5	9.3	447	30 <th>US-10-419-128-17422</th> <th>Sequence 17422, A</th>	US-10-419-128-17422	Sequence 17422, A

ALIGNMENTS

RESULT 1
US-10-049-473a-2
Sequence 2, Application US/10049473A
GENERAL INFORMATION:
APPLICANT: de Groot, Ronald
APPLICANT: Hermans, Peter Wilhelmus Maria
TITLE OF INVENTION: Pneumococcal Vaccines
FILE REFERENCE: Docket 294-120 PCT/US
CURRENT APPLICATION NUMBER: US/10/049, 473A
CURRENT FILING DATE: 2002-02-12
PRIOR APPLICATION NUMBER: PCT/NL00/00569
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: EP 99202640.1
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 322
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-10-049-473a-2
Query Match 100.0%; Score 1588; DB 26; Length 322;
Best Local Similarity 100.0%; Pred. No. 1,4e-113;
Matches 322; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 SNTYIKGVDMKKLLAGATITLLSVATLAACSGSGADLISMKGVITTEHOFYEQVKSNP 60
DB 1 SNTYIKGVDMKKLLAGATITLLSVATLAACSGSGADLISMKGVITTEHOFYEQVKSNP 60
QY 61 SAQOVLNMTIOKVEKQYSELDEKVDVDTIAEKKOYGENYQVLSQAGMTLETTRAAQ 120
DB 61 SAQOVLNMTIOKVEKQYSELDEKVDVDTIAEKKOYGENYQVLSQAGMTLETTRAAQ 120
QY 121 IRTSLVELAVKVAEELTDEAYKKAPEYTPDYTAQIIRLNNDKAEVLEKAKAGCA 180
DB 121 IRTSLVELAVKVAEELTDEAYKKAPEYTPDYTAQIIRLNNDKAEVLEKAKAGCA 180
QY 181 DEQAOKNSNDEKREKNGEITFDASSTEVEQYKKAFALDVGVSDVITATGTQAYS 240
DB 181 DEQAOKNSNDEKREKNGEITFDASSTEVEQYKKAFALDVGVSDVITATGTQAYS 240
QY 241 SOYIVKLTKEKSSNIDYKREKLTIVILQKQNDSTFVOSIIGKELOAANIKVKDQAF 300
DB 241 SOYIVKLTKEKSSNIDYKREKLTIVILQKQNDSTFVOSIIGKELOAANIKVKDQAF 300
QY 301 QNIFQYIGGGDSSSSSTSSNE 322
DB 301 QNIFQYIGGGDSSSSSTSSNE 322
```

RESULT 2

```
US-09-769-787-163
; Sequence 163, Application US/09769787
; GENERAL INFORMATION:
; APPLICANT: Microbial Techniques Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Phillip M
; FILE REFERENCE: PMC/P21129W
; CURRENT APPLICATION NUMBER: US/09/769,787
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 163
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-787-163
```

Query Match 97.0%; Score 1541; DB 22; Length 313;
Best Local Similarity 100.0%; Pred. No. 5,6e-110;
Matches 313; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 10 MKKLLAGATITLLSVATLAACSGSGADLISMKGVITTEHOFYEQVKSNP 69
DB 1 MKKLLAGATITLLSVATLAACSGSGADLISMKGVITTEHOFYEQVKSNP 69
QY 70 TIOKVEKQYSELDEKVDVDTIAEKKOYGENYQVLSQAGMTLETTRAAQ 129
DB 70 TIOKVEKQYSELDEKVDVDTIAEKKOYGENYQVLSQAGMTLETTRAAQ 129
QY 121 IRTSLVELAVKVAEELTDEAYKKAPEYTPDYTAQIIRLNNDKAEVLEKAKAGCA 180
DB 121 IRTSLVELAVKVAEELTDEAYKKAPEYTPDYTAQIIRLNNDKAEVLEKAKAGCA 180
QY 181 DEQAOKNSNDEKREKNGEITFDASSTEVEQYKKAFALDVGVSDVITATGTQAYS 240
DB 181 DEQAOKNSNDEKREKNGEITFDASSTEVEQYKKAFALDVGVSDVITATGTQAYS 240
QY 241 SOYIVKLTKEKSSNIDYKREKLTIVILQKQNDSTFVOSIIGKELOAANIKVKDQAF 300
DB 241 SOYIVKLTKEKSSNIDYKREKLTIVILQKQNDSTFVOSIIGKELOAANIKVKDQAF 300
```

```
QY 310 GGDSSSSSTSSNE 322
DB 301 GGDSSSSSTSSNE 313
```

RESULT 3

```
US-09-583-110-5200
; Sequence 5200, Application US/09583110
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: PAT00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 5200
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-5200
```

Query Match 96.9%; Score 1538; DB 19; Length 313;
Best Local Similarity 99.7%; Pred. No. 9,5e-110;
Matches 312; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
QY 10 MKKLLAGATITLLSVATLAACSGSGADLISMKGVITTEHOFYEQVKSNP 69
DB 1 MKKLLAGATITLLSVATLAACSGSGADLISMKGVITTEHOFYEQVKSNP 69
QY 70 TIOKVEKQYSELDEKVDVDTIAEKKOYGENYQVLSQAGMTLETTRAAQ 129
DB 70 TIOKVEKQYSELDEKVDVDTIAEKKOYGENYQVLSQAGMTLETTRAAQ 129
QY 121 IRTSLVELAVKVAEELTDEAYKKAPEYTPDYTAQIIRLNNDKAEVLEKAKAGCA 180
DB 121 IRTSLVELAVKVAEELTDEAYKKAPEYTPDYTAQIIRLNNDKAEVLEKAKAGCA 180
QY 181 DEQAOKNSNDEKREKNGEITFDASSTEVEQYKKAFALDVGVSDVITATGTQAYS 240
DB 181 DEQAOKNSNDEKREKNGEITFDASSTEVEQYKKAFALDVGVSDVITATGTQAYS 240
QY 241 SOYIVKLTKEKSSNIDYKREKLTIVILQKQNDSTFVOSIIGKELOAANIKVKDQAF 300
DB 241 SOYIVKLTKEKSSNIDYKREKLTIVILQKQNDSTFVOSIIGKELOAANIKVKDQAF 300
QY 310 GGDSSSSSTSSNE 322
DB 301 GGDSSSSSTSSNE 313
```

RESULT 4

```
US-09-107-433-4127
; Sequence 4127, Application US/09107433
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: Nucleic Acid and Amino Acid
; FILE REFERENCE: PAT00-07A
; CURRENT APPLICATION NUMBER: US/09/107,433
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5206
; SEQ ID NO 5206
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-107-433-4127
```

NUMBER OF SEQUENCES: 5206
CORRESPONDENCE ADDRESS:
ADDRESS: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354

```

COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: <Unknown>
OPERATING SYSTEM: <Unknown>
SOFTWARE: <Unknown>
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,433
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/ 085131
FILING DATE: May 12, 1998
APPLICATION NUMBER: 60/051553
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Arianello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 4127:
SEQUENCE CHARACTERISTICS:
LENGTH: 295 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEetical: YES
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: misc-feature
LOCATION: (B) LOCATION 1...295
SEQUENCE DESCRIPTION: SEQ ID NO: 4127:
US-09-107-433-4127

```

Query Match 89.9%; Score 1428; DB 15; Length 295;
 Best Local Similarity 99.3%; Pred. No. 2,6e-101;
 Matches 289; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

QY 32 KSESGADLISMKGDIYTHQFYEQYKSNPSAQOVLNMTIOKVFKEQYSGSELDKREVD 91
DB 5 KSESGADLISMKGDIYTHQFYEQYKSNPSAQOVLNMTIOKVFKEQYSGSELDKREVD 64
QY 92 IAEKKQYGENYQVRLSQAGMTLETRKAOIRTSKVELAVKVAEAEITDEAYKKAFDEY 151
DB 65 IAEKKQYGENYQVRLSQAGMTLETRKAOIRTSKVELAVKVAEAEITDEAYKKAFDEY 124
QY 152 TPDVTAQIIRLNNEDEKAEVLEKAKAEGADFAQLAKDNSTDEKTEKNGEITFDASASTEV 211
DB 125 TPDVTAQIIRLNNEDEKAEVLEKAKAEGADFAQLAKDNSTDEKTEKNGEITFDASASTEV 184
QY 212 PPOVKAALAVDGVSDYITATGQVASSOYIYKLTTRKTEKSSNIDYKREKLVIT 271
DB 185 PPOVKAALAVDGVSDYITATGQVASSOYIYKLTTRKTEKSSNIDYKREKLVIT 244
QY 272 QKQNSTFVQSIIGKELEKQANIKVKDQAFONIFTQYIGGSDSSSSSTSE 322
DB 245 QKQNSTFVQSIIGKELEKQANIKVKDQAFONIFTQYIGGSDSSSSSTSE 295

```

RESULT 5
 US-08-961-083-34
 Sequence 34, Application US/08961083
 GENERAL INFORMATION:

APPLICANT: Choi et. al.
 TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
 NUMBER OF SEQUENCES: 452
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA

```

ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE: 30-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: <Unknown>
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 213 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-08-961-083-34

```

Query Match 57.7%; Score 916; DB 13; Length 213;
 Best Local Similarity 93.6%; Pred. No. 4.5e-62;
 Matches 190; Conservative 0; Mismatches 5; Indels 8; Gaps 2;

```

QY 31 KSGSEADLISMKGDIYTHQFYEQYKSNPSAQOVLNMTIOKVFKEQYSGSELDKREVD 90
DB 1 KSGSEADLISMKGDIYTHQFYEQYKSNPSAQOVLNMTIOKVFKEQYSGSELDKREVD 60
QY 91 TIAEKKQYGENYQVRLSQAGMTLETRKAOIRTSKVELAVKVAEAEITDEAYKKAFDE 150
DB 61 TIAEKKQYGENYQVRLSQAGMTLETRKAOIRTSKVELAVKVAEAEITDEAYKKAFDE 120
QY 151 TTPDVTAQIIRLNNEDEKAEVLEKAKAEGADFAQLAKDNSTDEKTEKNGEITFDASASTE 210
DB 121 TTPDVTAQIIRLNNEDEKAEVLEKAKAEGADFAQLAKDNSTDEKTEKNGEITFDASASTE 180
QY 211 VP-EQYKAAFA-----LDVD 225
DB 181 VPGASPKPLFAFCGMVFLDVD 203

```

RESULT 6
 US-09-765-271-34
 Sequence 34, Application US/09765271
 GENERAL INFORMATION:

APPLICANT: Choi et. al.
 TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
 NUMBER OF SEQUENCES: 452
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA
 ZIP: 20850
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/765,271
 FILING DATE: 22-Jan-2001


```

: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: SEQUENCE DESCRIPTION: SEQ ID NO: 34
US-09-765-272A-34

```

Query Match	916:	57.7%:	Score	916:	DB	22:	Length	213;
Best Local Similarity		93.6%:	Pred. No.	4.5e-62:				
Matches	190;	Conservative	0;	Mismatches	5;		Indels	8;
							Gaps	2

Qy	31	SKSGSGADLISMKGVITEHOFYEQVKSMP\$QOVLNNT	IQKVEFKQYSGELDDXEVD	90
Db	1	SKSGSGADLISMKGVITEHOFYEQVKSMP\$QOVLNNT	IQKVEFKQYSGELDDXEVD	60
Qy	91	1TIAEKKOYGENYOQVLSOAGMTLETTRKAOIFTSKIVELAVKVAEAEITDEAYKKAPE	1500	
Db	61	1TIAEKKOYGENYOQVLSOAGMTLETTRKAOIFTSKIVELAVKVAEAEITDEAYKKAPE	1200	
Qy	151	YTPDVTAOIIRLNNEDKAEVLEKKAABGADRAQALAKNDSTDEKTEKNGEITPDSASTE	210	
Db	121	YTPDVTAOIIRLNNEDKAEVLEKKAABGADRAQALAKNDSTDEKTEKNGEITPDSASTE	1800	
Qy	211	VP-EQVKKAAFA-----LDVD	225	
Db	181	VPGASPKKPLFAFCGMWFLDVD	203	

RESULT 9

US-10-417-884-6355
; Sequence 6355, Application US/10417884

APPLICANT: Lynn A Doucette-Stamm and David Bush
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 ENTEROCOCCUS FACIUM FOR DIAGNOSTICS AND THERAPEUTICS
 NUMBER OF SEQUENCES: 7310

ADDRESS: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA

```
;;  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII
```

APPLICATION NUMBER: US/10/417,884
FILING DATE: 17-Apr-2003

APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998

FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571

ATTORNEY/AGENT INFORMATION:
NAME: Arinello, Pamela Deneke

REFERENCE/DOCKET NUMBER: GTC-0122
TELECOMMUNICATION INFORMATION:

TELEPHONE: (781) 893-8277
TELEFAX: (781) 893-8277
INFORMATION FOR SEO ID NO: 6355

```

?
? SEQUENCE CHARACTERISTICS:
? LENGTH: 344 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? HYPOTHEetical: YES
? ORIGINAL SOURCE:
? ORGANISM: Enterococcus faecium

```

```

;      FEATURE:
;      NAME/KEY: misc feature
;      LOCATION: (B) LOCATION 1...344
;      SEQUENCE DESCRIPTION: SEQ ID NO: 6355.
;
US-10-417-884-6355
```

Query Match	39.9%	Score 634	DB 30	Length 344
Best Local Similarity	40.5%	Pred. No. 5e-40		
Matches 130	Conservative 75	Mismatches 106	Indels 10	Gaps 5

[illegible]

RESULT 10

US-09-134-000-4284
; Sequence 4284, Application US/09134000A

APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS

```

; FILE REFERENCE: GTC-005
; CURRENT APPLICATION NUMBER: US/09/134,000A
; CURRENT FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 6810
; SEQ ID NO 4284

```

Query Match	38.3%	Score 608.5	DB 15	Length 368
Best Local Similarity	41.8%	Pred. No. 5e-38		
Matches 132	Conservative 63	Mismatches 112	Indels 9	Gaps 5

Qy	10	HKKLLIAGATLLSVATLAACSKSGEADLISMGDVITHEQFEYOKSNPSAQOVLNM	69
		: : : : : : : : : : :	
Db	27	MKKLLIILAAGAAVFSLAACSSGSK - DIATMGSTITVDFFFNQIKRESTSOQAPSOM	84
		: : : : : : : : : : :	
Oy	70	TIOVFVEKQYSESLDDKEVDITIAEKKO - -YGENYORVLSQAGMLLETRKAQIRSKL	126
		: : : : : : : : : : :	
Db	85	VIVYFVEEKYDXTDKDIOKNPDEAKQVYVAGSGKSPSDALKQAGLTKETPKKOLGRRA	144
		: : : : : : : : : : :	
Oy	127	VELAVKAAEALTDEAVKKAFFDEYTPDVTQAOTRLNDEKAKVELEKAKAEGADPAOLA	186
		: : : : : : : : : : :	
Db	145	YDAGLTK - AHIKLTDEDELKTAMASFHPEVEAOIIQVASEDDAKAV - KKETIDGGDFPKYA	201
		: : : : : : : : : : :	
Oy	187	KDNSTDEKTEGGEITPDASTASTPEBQVKKAAAFALVDGSDVITPTGTQAYSSOYIYI	246
		: : : : : : : : : : :	
Db	202	KEKSTDTATTKKDGSKIFKDSQATITVPAYEVKDLAAKRLKDGSEPSFIATIMQYITTYIYV	261
		: : : : : : : : : : :	
Oy	247	KLTKTEKSSNIDYKEKLTIVLTQKONDSFVOSIIGELQQAANIKVYKDAQONIFQ	306
		: : : : : : : : : : :	

Fri Sep 5 09:24:50 2003

us-10-049-473a-2.rapm

Page 6

Db 262 KMTNKKAKGNDMPYEKEIKIAETKRLADQTFVSVISDELKAAVNIKIDDAFKNALAG 321
Y 307 YIGGDSSTSSSTSSNE 322
Db 322 YM-QTESSASSEKKE 336

RESULT 11
US-09-134-000C-4284
Sequence 4284, Application US/09134000C
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
FILE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 032796-032
CURRENT FILING DATE: 1998-08-13
CURRENT APPLICATION NUMBER: US 60/055,778
PRIOR FILING DATE: 1997-08-15
NUMBER OF SEQ ID NOS: 6812
SOFTWARE: PatentIn version 3.1
SEQ ID NO 4284
LENGTH: 368
TYPE: PRN
ORGANISM: Enterococcus faecalis
US-09-134-000C-4284

Query Match 38.3%; Score 608.5; DB 15; Length 368;
Best Local Similarity 41.8%; Pred. No. 5e-38;
Matches 132; Conservative 63; Mismatches 112; Indels 9; Gaps 5;
Y 10 MKKILAGATITLLSVATLAACSGSEGADLISMKGDVITEHQFVEQVKNPSAQVLLNM 69
Db 27 MKKILILAAAGAMVFLAACSSGSK--DIATMKGSTIIVDDFYNGIKQSTSQQAFSQM 84
Y 70 TIOKVEKQVSELDDEVDITIAEKKQ--YGENYQVLSQAGMTLETTRKAQIRTSKL 126
Db 85 VIKVFEKQVSELDDEVDITIAEKKQ--YGENYQVLSQAGMTLETTRKAQIRTSKL 144
Y 127 VELAARKVAEALDEAVKKAFADEYTPDVTAQIIRLNNEDEKAEVLEKAKAGADFAOLA 186
Db 145 YDAGLK--AHLKTTDEDLKTAMASFHPEVQAQIIOVASEDDAKAV--KKEITDGDFTKIA 201
Y 187 KDNSTDEKTEKNGEITFDSASTVEPEQVKAFAALDVGVSDVITATGTQAVSSQYIY 246
Db 202 KESTDTATKKGDKIKFDSQATVPAVEKAEAFKLDGEVSEPIAATNMOTYTYYV 261
Y 247 KTKKTEKSSNIDYKREKIKTYILTKQKNDSTFVOSIIGKELOANIKVKAQAFONITQ 306
Db 262 KMTNKKAKGNDMPYEKEIKIAETKRLADQTFVSVISDELKAAVNIKIDDAFKNALAG 321
Y 307 YIGGDSSTSSSTSSNE 322
Db 322 YM-QTESSASSEKKE 336

RESULT 12
US-10-434-665-4284
Sequence 4284, Application US/10434665
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
FILE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: PATH03-09
CURRENT APPLICATION NUMBER: US/10/434,665
CURRENT FILING DATE: 2003-05-14
PRIOR APPLICATION NUMBER: US 09/134,000
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/055,778
PRIOR FILING DATE: 1997-08-15
NUMBER OF SEQ ID NOS: 6812
SOFTWARE: PatentIn version 3.1

SEQ ID NO 4284
LENGTH: 368
TYPE: PRN
ORGANISM: Enterococcus faecalis
US-10-434-665-4284

Query Match 38.3%; Score 608.5; DB 30; Length 368;
Best Local Similarity 41.8%; Pred. No. 5e-38;
Matches 132; Conservative 63; Mismatches 112; Indels 9; Gaps 5;
Y 10 MKKILAGATITLLSVATLAACSGSEGADLISMKGDVITEHQFVEQVKNPSAQVLLNM 69
Db 27 MKKILILAAAGAMVFLAACSSGSK--DIATMKGSTIIVDDFYNGIKQSTSQQAFSQM 84
Y 70 TIOKVEKQVSELDDEVDITIAEKKQ--YGENYQVLSQAGMTLETTRKAQIRTSKL 126
Db 85 VIKVFEKQVSELDDEVDITIAEKKQ--YGENYQVLSQAGMTLETTRKAQIRTSKL 144
Y 127 VELAARKVAEALDEAVKKAFADEYTPDVTAQIIRLNNEDEKAEVLEKAKAGADFAOLA 186
Db 145 YDAGLK--AHLKTTDEDLKTAMASFHPEVQAQIIOVASEDDAKAV--KKEITDGDFTKIA 201
Y 187 KDNSTDEKTEKNGEITFDSASTVEPEQVKAFAALDVGVSDVITATGTQAVSSQYIY 246
Db 202 KESTDTATKKGDKIKFDSQATVPAVEKAEAFKLDGEVSEPIAATNMOTYTYYV 261
Y 247 KTKKTEKSSNIDYKREKIKTYILTKQKNDSTFVOSIIGKELOANIKVKAQAFONITQ 306
Db 262 KMTNKKAKGNDMPYEKEIKIAETKRLADQTFVSVISDELKAAVNIKIDDAFKNALAG 321
Y 307 YIGGDSSTSSSTSSNE 322
Db 322 YM-QTESSASSEKKE 336

RESULT 13
US-10-206-576-182
Sequence 182, Application US/10206576
GENERAL INFORMATION:
APPLICANT: Choi et al.
TITLE OF INVENTION: Enterococcus faecalis polynucleotides and polypeptides
NUMBER OF SEQUENCES: 497
CORRESPONDENCE ADDRESS:
ADDRESS: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-R
COMPUTER: Dell Latitude
OPERATING SYSTEM: Windows 98
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/206,576
FILING DATE: 29-Jul-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/071,035
FILING DATE: 1998-05-04
APPLICATION NUMBER: US 60/046,655
FILING DATE: 1997-05-16
APPLICATION NUMBER: US 60/044,031
FILING DATE: 1997-05-06
APPLICATION NUMBER: US 60/066,009
FILING DATE: 1997-11-14
ATTORNEY/AGENT INFORMATION:
NAME: Hyman, Mark J.
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB369P1D1
INFORMATION FOR SEQ ID NO: 182:
SEQUENCE CHARACTERISTICS:

Fri Sep 5 09:24:50 2003

APPLICATION NUMBER: US 60/044,031
FILING DATE: 1997-05-06
APPLICATION NUMBER: US 60/066,009
FILING DATE: 1997-11-14
ATTORNEY/AGENT INFORMATION:
NAME: Hyman, Mark J.
REGISTRATION/DOCKET NUMBER: PB369P1D1
REFERENCE/DOCKET NUMBER: PB369P1D1
INFORMATION FOR SEQ ID NO: 184:
SEQUENCE CHARACTERISTICS:
LENGTH: 322 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 184:
US-10-206-576-184

Query Match 35.4%; Score 562.5; DB 28; Length 322;
Best Local Similarity 41.2%; Pred. No. 1,5e-34;
Matches 122; Conservative 58; Mismatches 107; Indels 9; Gaps 5;
30 CSNGSGADLISMKGSDVITEHQEYQVKSQFSAQVILNMTIQVFEKQYGSLEDDKEVD 89
1 CSSGSK--DIATMKGSTIVDDFYNOIKQOSTSOQAFSQMVIYKVFEEKYGDKVDKXIO 58
90 DTIAEKKQ---YGENYQVLSQAGMTLETRKRAQIRTSKIVELAVKVAEAEILTDEAYKK 146
59 KNFDEAKEQVEAOGKGFSDALMKAQGLTERKFKQIKORAAVDAGLK--AHKLTDEDLKT 116
147 AFDEYTPDYTAQIIRLNNEDEKAEVLEKAKAGADFAQLANDNSTDEKTEKENGSETPDS 206
117 AMASHPPEVEKAOIIQVASEDDAKAV--KKEITDGGDFTKAKKSTDTATKKGKIKEDS 175
207 ASTEYPEQVKKAFALDVDSVITATGTQAVSSQYIVKLTKEKSSNIDYKEXLK 266
176 QATVPAEYKKAFLKNGEYSEPIAATNMQTYYVVKTKKAKGNDMKPYEKEIK 235
267 TVILTKQKNDSTFVQSTIGKELQANIKVRDOAFONITQYIGGDDSSSSSTSN 322
236 KIAEETKRLADQTFVSKVISDELKAAVYKIKDDAFKNALAGYM-QTESSSSASSEKKE 290

Search completed: August 29, 2003, 10:44:00
Job time : 380 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2003, 10:34:29 ; Search time 25 Seconds
(Without alignments)
381.116 Million cell updates/sec

Title: US-10-049-473a-2
Perfect score: 1588
Sequence: 1 SNTYLGVDPMKKILAGAIR.....IFTQYIGGDSSTSSSTSSNE 322

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141812 seqs, 29589763 residues

Total number of hits satisfying chosen parameters: 141812

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Pending_Patents_AA_New: *
1: /cgn2_6/ptodata/2/paa/PCT_NEW.COMB.pep: *
2: /cgn2_6/ptodata/2/paa/US06_NEW.COMB.pep: *
3: /cgn2_6/ptodata/2/paa/US07_NEW.COMB.pep: *
4: /cgn2_6/ptodata/2/paa/US08_NEW.COMB.pep: *
5: /cgn2_6/ptodata/2/paa/US09_NEW.COMB.pep: *
6: /cgn2_6/ptodata/2/paa/US10_NEW.COMB.pep: *
7: /cgn2_6/ptodata/2/paa/US60_NEW.COMB.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB ID	Description
1	1428	89.9	295	US-10-617-320-4127	Sequence 4127, Ap
2	163	10.3	621	US-09-897-516A-7078	Sequence 7078, Ap
3	129.5	8.2	589	US-10-299-636-97	Sequence 97, Appl
4	129.5	8.2	929	US-10-299-636-94	Sequence 94, Appl
5	129.5	8.2	1314	US-60-487-610-2694	Sequence 2694, Ap
6	127.5	8.0	1790	US-10-326-956-1291	Sequence 1291, Ap
7	126	7.9	643	US-10-299-636-95	Sequence 95, Appl
8	123.5	7.8	630	US-10-603-113-20275	Sequence 20275, A
9	123	7.7	588	US-10-299-636-96	Sequence 96, Appl
10	122	7.7	901	US-10-603-113-14747	Sequence 14747, A
11	121.5	7.7	802	US-10-273-573-9033	Sequence 9033, Ap
12	121.5	7.7	1154	US-10-273-573-9040	Sequence 9040, Ap
13	119	7.5	336	US-10-299-636-103	Sequence 103, App
14	119	7.5	354	US-10-299-636-105	Sequence 105, App
15	119	7.5	713	US-10-257-377-3	Sequence 3, Appl
16	118.5	7.5	2101	US-09-296-662B-32	Sequence 32, Appl
17	118.5	7.5	2207	US-10-273-573-9039	Sequence 9039, Ap
18	117.5	7.4	200	US-10-299-636-50	Sequence 50, Appl
19	117.5	7.4	2115	US-09-296-662B-33	Sequence 33, Appl
20	117.5	7.4	2121	US-10-273-573-9034	Sequence 9034, Ap
21	116.5	7.3	301	US-10-603-113-26282	Sequence 26282, A
22	116.5	7.3	1875	US-10-326-956-2369	Sequence 2369, Ap
23	116	7.3	979	US-10-603-114-5466	Sequence 5466, Ap
24	116	7.3	1014	PCT-US03-21379-81	Sequence 81, Appl
25	116	7.3	1960	US-10-293-244-1516	Sequence 1516, Ap
26	116	7.3	1963	US-10-293-244-3484	Sequence 3484, Ap

27	115.5	7.3	299	US-09-065-756B-61	Sequence 61, Appl
28	115.5	7.3	448	US-10-603-113-20279	Sequence 20279, A
29	115.5	7.3	635	US-10-603-114-5956	Sequence 5956, Ap
30	115	7.2	1988	US-10-286-897-5930	Sequence 5930, Ap
31	115	7.2	1988	US-10-286-897-5931	Sequence 5931, Ap
32	115	7.2	1988	US-10-258-898A-5930	Sequence 5930, Ap
33	115	7.2	1988	US-10-258-898A-5931	Sequence 5931, Ap
34	114.5	7.2	2117	US-10-273-573-9038	Sequence 9038, Ap
35	114	7.2	639	US-10-603-113-20583	Sequence 20583, A
36	112.5	7.1	206	US-10-612-783-5800	Sequence 5800, Ap
37	112.5	7.1	1341	US-10-603-113-19074	Sequence 19074, A
38	111.5	7.0	487	US-10-333-120A-9	Sequence 9, Appl1
39	111	7.0	428	PCT-US02-34769-8	Sequence 8, Appl1
40	111	7.0	429	PCT-US02-34769-16	Sequence 16, Appl
41	111	7.0	1213	US-10-286-897-3161	Sequence 3161, Ap
42	111	7.0	1213	US-10-258-898A-3161	Sequence 3161, Ap
43	110.5	7.0	299	US-09-065-756B-341	Sequence 341, App
44	110.5	7.0	417	US-60-487-610-1535	Sequence 1535, Ap
45	110.5	7.0	427	US-10-326-956-2877	Sequence 2877, Ap

ALIGNMENTS

RESULT 1
US-10-617-320-4127
Sequence 4127, Application US/10617320
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NOCLETIC ACID AND AMINO ACID
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIA
THERAPEUTICS
NUMBER OF SEQUENCES: 5206
CORRESPONDENCE ADDRESSES:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: <Unknown>
OPERATING SYSTEM: <Unknown>
SOFTWARE: <Unknown>
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/617,320
FILING DATE: 10-Jul-2003
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/107,433
FILING DATE: 30-Jun-1998
APPLICATION NUMBER: 60/085131
FILING DATE: May 12, 1998
APPLICATION NUMBER: 60/051553
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Artinello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 4127:
SEQUENCE CHARACTERISTICS:
LENGTH: 295 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: misc_feature

LOCATION: (B) LOCATION 1...295
SEQUENCE DESCRIPTION: SEQ ID NO: 4127:
US-10-617-320-4127

Query Match 89.9%; Score 1428; DB 6; Length 295;
Best Local Similarity 99.3%; Pred. No. 6e-93;
Matches 289; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 32 KSGGALLIMKGVITTEHOFEYEVKNSPSAQOVLMMTIOKVEKQYSGELDDKEDDT 91
DB 5 KVSAGALLIMKGVITTEHOFEYEVKNSPSAQOVLMMTIOKVEKQYSGELDDKEDDT 64
QY 92 IAEKKQYGENYQVLSQAGMTLETTRKAQIRTSKLVELAVKVAEALTEBAKKARDEY 151
DB 65 IAEKKQYGENYQVLSQAGMTLETTRKAQIRTSKLVELAVKVAEALTEBAKKARDEY 124
QY 152 TPDTYAOIIRLNNDKAKVELEKAKAGADFAQALAKNSNDEKTEKNGGETTDSASTEV 211
DB 125 TPDTYAOIIRLNNDKAKVELEKAKAGADFAQALAKNSNDEKTEKNGGETTDSASTEV 184
QY 212 PEQYKKAFAFDVGVSDVITATGTQAVSSQYIVKLTKEKSSNIDYKREKLTIVLT 271
DB 185 PEQYKKAFAFDVGVSDVITATGTQAVSSQYIVKLTKEKSSNIDYKREKLTIVLT 244
QY 272 QKQNDSTFVQSIIGKELQAAANIKVKDAFONITFYIGGDDSSSSSTSNE 322
DB 245 QKQNDSTFVQSIIGKELQAAANIKVKDAFONITFYIGGDDSSSSSTSNE 295

RESULT 2

US-09-897-516a-7078

Sequence 7078, Application US/0989/516A

GENERAL INFORMATION:

APPLICANT: Corbin, David R.
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Huesing, Joseph E.
APPLICANT: Malvar, Thomas M.
APPLICANT: Krasomil-Osterfeld, Karina C.
APPLICANT: Slater, Steven C.
APPLICANT: Spiridonov, Sergei
TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
FILE REFERENCE: 38-21(51847)B
CURRENT APPLICATION NUMBER: US/09/897, 516A
CURRENT FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/215,161
PRIOR FILING DATE: 2000-06-30
NUMBER OF SEQ ID NOS: 8415
SEQ ID NO 7078
LENGTH: 621
TYPE: PRT
ORGANISM: Xenorhabdus sp.
US-09-897-516a-7078

Query Match 10.3%; Score 163; DB 5; Length 621;
Best Local Similarity 20.4%; Pred. No. 0.00042;
Matches 92; Conservative 72; Mismatches 136; Indels 152; Gaps 19;

QY 13 KLGAATLLSVATLAACSGSEGADLIS--MKGVITE--HQFEYQV----- 57
DB 15 KIVALLILITFLVGLTGYLSSEGNYAAKNGCQISRAQLEQVFOQKMSLOERLDOF 74
QY 58 -----SNPSAQOVL-----LNMIOKVEKQYSGELDDKEDDTIAEE--KKQY----- 99
DB 75 SALLSNEQEVRLAKROSIDSIIINTVLLLEOYANKLGLSASDQVKEIRKQRYFOTDGF 134
QY 100 GENYQVLSQAGMTLETTRKAQI-----RTSK 125
DB 135 NEKTIQVITNAGLNDNNAEQIRQDLISROLRTILMSTELPAEIKQALQSLQEKRRAR 194
QY 126 LVELAVKRV-AEALTEDEA----- 143
DB 195 LATLELKSTIEAQOVSVDQELKKNYMNMSKHTFVEKVKISYIKMDADELKNVTVDADI 254

QY 144 ---YKKADEYTPDVTAO---IIRLNNDKAKVELEKAKAGADFAQALAKNSNDEKTEK 198
DB 255 EKTYKNNLSKTYTPPEKKKYSLLIOLASEADAKSVLDELK-KGADFSKLAEEKSTDYKSAQ 313
QY 199 GGEITPDSASTEVPEQYKKAFAFDVGVSDVITATGTQAVSSQYIVKLTKEKSSN 257
DB 314 GGLGWMNEESS--PDEL-KSAHLTEKQOLSAVI-----KLSNGYAVFRLDIDIKPQVAP 365
QY 258 IDDKKELKTVILITQKQNDSTF-----OSIIGKELQAAANIKV----- 295
DB 366 LADERSAIEKTVQOEALIDFVALQRRKVSSEAAANDESLAAE-QAAGTKAVTTDMFDD 424
QY 296 ---KQDAFONITFYIGG--DSSSSSTSNE 322
DB 425 HVPADINFQVQVAFIFGNLVDDKSPSGTNSD 456

RESULT 3

US-10-299-636-97

Sequence 97, Application US/10299636

GENERAL INFORMATION:

APPLICANT: Briles, David E.
APPLICANT: McDaniel, Larry S.
APPLICANT: Swatello, Edwin
APPLICANT: Yohef, Janet
APPLICANT: Crain, Marilyn J.
APPLICANT: Hollingshead, Susan
APPLICANT: Tart, Rebecca
TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEINS AND USES THEREOF
FILE REFERENCE: 57909/361
CURRENT APPLICATION NUMBER: US/10/299, 636
CURRENT FILING DATE: 2002-11-19
PRIOR APPLICATION NUMBER: 08/714,741
PRIOR FILING DATE: 1996-09-16
PRIOR APPLICATION NUMBER: 08/529,055
PRIOR FILING DATE: 1995-09-15
NUMBER OF SEQ ID NOS: 111
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 97
LENGTH: 589
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-10-299-636-97

Query Match 8.2%; Score 129.5; DB 6; Length 589;
Best Local Similarity 22.5%; Pred. No. 0.087;
Matches 82; Conservative 53; Mismatches 125; Indels 105; Gaps 16;

QY 42 MKGVITEHOFEYEVKNSPSA--QOVL-----LNMIOKVEKQYSGELDDKEDDTIAEE 95
DB 232 VSGELATPDKRENDKSSDSSVGEETLPSPLNNANSQTEHR-----KQVDEYI--- 281
QY 96 KKQYGE--NYQVLSQAGMTLETTRKAQIRTSKLVELAV----- 131
DB 282 KKMSELQDLRRKHTQOVNINIKISAIKTYIVELSVLKNSKKEELTSTKALTEFAFE 341
QY 132 -----KKAFAE-LTDEAYKKADEYTPD----- 154
DB 342 QFKDQTLKPEKVAEAKKVEAKKAKKADKEDRRNYPNTYKTTLELTAESDVYKKA 401
QY 155 -----VTQOIRLNNDKAKVELEKAKAGADFAQALAKNSNDEKTEKNGGETTDSASTE 210
DB 402 ELELVKEANESREERIKQAKKEVESKKAATRLER-ITDRKKA-----EAKRK 453
QY 211 VPEQYKKAFAFD--VDSVSDVITATGTQAVSSQYIVKLTKEK--SSNIDY--KEKLR 266
DB 454 AESEKKAAREKQKVDAEVALAE--KIALEVEYORLEKELKEIDSESDYLKGLR 510
QY 267 TVI---ITQKQNDSTFVQSIIGK-----ELQAAANIKVKDAFONITFYIGGDDSSSS 317
DB 511 APLQSKIDTYKAKLSKLEELSDKIDELDAEIAKLEVLQKDAEGNNNVEAYFEGLEKETA 570

QY 318 STSNE 322
Db 571 EKRAE 575

RESULT 4

US-10-299-636-94
; Sequence 94, Application US/10299636
; GENERAL INFORMATION:
; APPLICANT: Billes, David E
; APPLICANT: McDaniel, Larry S
; APPLICANT: Swatlo, Edwin
; APPLICANT: Yotter, Janet
; APPLICANT: Crain, Marilyn J
; APPLICANT: Hollingshead, Susan
; APPLICANT: Tarr, Rebecca
; APPLICANT: Brooks-Walter, Alexis
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEINS AND USES THEREOF
; FILE REFERENCE: 57909/361
; CURRENT APPLICATION NUMBER: US/10/299,636
; PRIOR FILING DATE: 2002-11-19
; PRIOR APPLICATION NUMBER: 08/714,741
; PRIOR FILING DATE: 1996-09-16
; PRIOR APPLICATION NUMBER: 08/529,055
; PRIOR FILING DATE: 1995-09-15
; NUMBER OF SEQ ID NOS: 111
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 94
; LENGTH: 929
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-299-636-94

Query Match 8.2%; Score 129.5; DB 6; Length 929;
Best Local Similarity 22.5%; Pred. No. 0.15;

Matches 82; Conservative 53; Mismatches 125; Indels 105; Gaps 16;

QY 42 MKGDVITEHOFYEQVKNPSA--QOVL---LNMFTQKVEKQVGSLEDDKEVDYTAEE 95
Db 269 VSGELATPPKKENDAKSSDSSVGEETLPSPLNMANESQTEHR-----KDVDEYI--- 318
QY 96 KQYGE-NYQVRLSQAGMTLETRKAQIRTSKLVELAY----- 131
Db 319 KKMSEIQDRRKHTQVNNIKLSAIKTYIVELSLKNSKKEELSTKALTAFAE 378
QY 132 -----KKVAEAE-LTDEAYKKAPEDEYTPD----- 154
Db 379 QFKKDTLPKPKYVAEAEKKVEAKKAKKQKEDRRNYPNTYKYLELTAESDVKYEA 438
QY 155 -----VTAQIIRLNNEDEAKVELEKAKAGDFQNLADNSTDEKTEKNGEITPDSASTE 210
Db 439 ELEIVKEANESRREERIKQAKKEVESKKAETRLER-ITDRKKAEE-----EEAKRK 490
QY 211 VPEQVKKAFALD-VDSVSDVITATGQAVSSQYIYIKTRKTEK--SSNIDY-KKKLK 266
Db 491 AESEKKAARAKQKVDAEYALEA---KIAELEVEYGRLEKELKEIDESDSEDLKGLR 547
QY 267 TVI---LTQKONDSFVQSIIGK-----ELQANIKVQDAFONITFYIYGGSSSSS 317
Db 548 APLQSKLDTRKAKLSKLEELSDKIDELDAETIAKLEVLKDAEGNNNEAVFKKELEKTTA 607
QY 318 STSNE 322
Db 608 EKRAE 612

RESULT 5

US-60-487-610-2694
; Sequence 2694, Application US/60487610
; GENERAL INFORMATION:
; APPLICANT: CARCILL, Michele
; APPLICANT: HUANG, Hongjin

; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: LIVER FIBROSIS IN HEPATITIS C VIRUS-INFECTED SUBJECTS,
; FILE REFERENCE: C0001469
; CURRENT APPLICATION NUMBER: US/60/487,610
; CURRENT FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 97101
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2694
; LENGTH: 1314
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-487-610-2694

Query Match 8.2%; Score 129.5; DB 7; Length 1314;
Best Local Similarity 20.9%; Pred. No. 0.24;
Matches 65; Conservative 73; Mismatches 136; Indels 37; Gaps 13;

QY 32 KSGEADLISMKG---DVITEHOFYEQVKNPSAQVLLNMFTQKVEKQVGSLEDDK-- 86
Db 753 KQGELESLINLGTGDEMVTQDAFEDRLK---EKEREYMDLLRAQOVKRDVQNTLMRLQ 809
QY 87 EVDDTTAEKKQYGENYQVLSQAGMTLETRKAQI-RTSKLVELAVKKVAEELTDEAYK 145
Db 810 RVNNTLSSQISRL-QNIRNTIEETGNLAQARAHVENTERLTIASRELEKAKVAANVS 868
QY 146 KAPEYTPDYTAQIIRLNN-----EDRAKEVLEKAKAGAPQALAK---DNSTDE----- 193
Db 869 VTQPESTGDP-----NNMTLLAEERKLAERIKQEDADIVRAKANDSTTEAVYILL 921
QY 194 KTKENGEITFPD-SASTEPEQVKKAFALDVG--VSDVITATGTOA---YSSQYIYK 247
Db 922 RTLAGENQTAPEIETELRKKEEQAKNISQDEKQAARHEAKKAGDAVELIYASVQDSP 981
QY 248 LTKRT--EKSNNIDYREKLTQVILITQKONDSFVQSIIGKELQANIKVQDAFONITF 305
Db 982 LDSETELENNIKMEAEINLEQLIDKLYEDLRDMRKELEVKMLKEKTEQOOTAD 1041
QY 306 QYIGGGSSSS 316
Db 1042 QILARADAARA 1052

RESULT 6

US-10-326-956-1291
; Sequence 1291, Application US/10326956
; GENERAL INFORMATION:
; APPLICANT: Bauer et al.
; TITLE OF INVENTION: Protein Complexes and Methods for their use
; FILE REFERENCE: 220615
; CURRENT APPLICATION NUMBER: US/10/326,956
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: EP 01 130 253.6
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 3282
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1291
; LENGTH: 1790
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-326-956-1291

Query Match 8.0%; Score 127.5; DB 6; Length 1790;
Best Local Similarity 22.8%; Pred. No. 0.48;
Matches 69; Conservative 59; Mismatches 103; Indels 71; Gaps 15;

QY 49 EHOFEYEQV---SNPSAQOVLNMTQKVEKQY-----SELDKEVD--D 90
Db 1172 ERQYNEDISQUNDITSTQGE--NESIKKKNDLEGEVKKMKSTSEQSNLKKSEIDALN 1229
QY 91 TIAEEKQYGENYQVLSQAGMTLETRKAQI-----RTSKLVELAVK----- 132
Db 1230 LQIKELKKKKTETNASLSEISIKVSESETVAKIQLQEDCNFKKEKVESELEKLRKASEDKNS 1289

```

QY      133  KVAEAELEDEYKKAFFEDYFDYVAQIIRLNNDEKAKVLEKARAEGBDFOLAKONSTD 192
Db      1290  KYLELOKRESEIKELBDKTELKIQLEKITNLSKAE-----KSE-SELSRLKTSSEE 1343
QY      193  EKTENGGEIITFDSASTEVEPOVKAAAFALDVGDSVITATGTQOAVSSOX-----YIVK 247
Db      1344  RKNAAE-----QLEKLKNKEI--QINQAFEREKRLKNLEG-SSTIQEISEKINTIIEDELIR 1396
QY      248  LTKRTE-KSSNIDYKERKLTV-----ILTQKONDSFVOSITGKELQANIRKVDQAF 300
Db      1397  LQNNENELKAEIDNTRSELEKVSLSNDELLEEKONTI-----KSLQDEILSYKKKIT 1448
QY      301  ON 302
Db      1449  RN 1450

```

```

RESULT 7
US-10-299-636-95
: Sequence 95, Application US/10299636
: GENERAL INFORMATION:
: APPLICANT: Briles, David E
: APPLICANT: McDaniel, Larry S
: APPLICANT: Swiatlo, Edwin
: APPLICANT: Yoether, Janet
: APPLICANT: Crain, Marilyn J
: APPLICANT: Hollingshead, Susan
: APPLICANT: Tart, Rebecca
: APPLICANT: Brooks-Walter, Alexis
: TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEINS AND USES THEREOF
: FILE REFERENCE: 57909/361
: CURRENT APPLICATION NUMBER: US/10/299,636
: CURRENT FILING DATE: 2002-11-19
: PRIOR APPLICATION NUMBER: 08/714,741
: PRIOR FILING DATE: 1996-09-16
: PRIOR APPLICATION NUMBER: 08/529,055
: PRIOR FILING DATE: 1995-09-15
: NUMBER OF SEQ ID NOS: 111
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 95
: LENGTH: 643
: TYPE: prt
: ORGANISM: Streptococcus pneumoniae
: US-10-299-636-95

```

Query Match	7.9%;	Score 126;	DB 6;	Length 643;
Best Local Similarity	22.8%;	Pred. No. 0.17;		
Matches 79;	Conservative 47;	Mismatches 117;	Indels 104;	Gaps 15;

```

QY 5E0VKNPNSAOQVLLMMTIOKEKQYGESELDDEKEDVDIIEAEKKOYQG-NYQVLSQAGM 112
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7 EETLSPSPS-----LNMANESQTEHR-----KQVDEYI-----KMLSLIOJDRKHHQNV 51
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 113 TETPKRAQRTSKLWEVLAV-----KQVAAE- 138
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 52 NINILSLAIKITYLLEISVLKENSKEELTSTKAELTAPEQPKDTLAPPEKVAEAER 111
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 139 -LTDEAYKKAFADEYTPD-----YTAQIIRLNNECKA 168
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 112 KVEEAKKKAQKQKEDDRNRNYPNTYKTELEIAESDVKFEAELELVKEANESRNEKI 171
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 169 KEVLKAKAEGADFOALAKDNSTDEKTEKENGELTFPSASIEVEYOYKKAFAALD-VDGV 227
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 172 KOAKKRVSSKKAERARLER-IKTDDKKA-----EAKKKAEESEKKAERKQVDAE 223
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 228 SDVITATQOAVSSOYIVLKLTKTEK-----SSNIDY-KEKLITYI---LQOKDNSTFYQ 281
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 224 EYALAE---KIALEIEYQRLKELEKLEIDESDSYLKEGRAPALQSKLDOTTKAKLSKE 280
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 282 SLICK-----ELQAAINKYVDQAFQNFIOYIGGGSSSSSSSTSNE 322
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 261 EISDKIDELDAEIAKLEVOQLDABGNNNVVEYFEQLEKTTAAEKAE 327
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 8
US-10-603-113-20275
: Sequence 20275, Application US/10603113
: GENERAL INFORMATION:
: APPLICANT: Keith Weinstock et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBI
: TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.132
: CURRENT APPLICATION NUMBER: US/10/603.113
: CURRENT FILING DATE: 2003-06-24
: PRIOR APPLICATION NUMBER: US/09/248,796
: PRIOR FILING DATE: 1999-02-12
: NUMBER OF SEQ. ID NOS: 28206
: SEQ ID NO 20275
: LENGTH: 630
: TYPE: PRN
: ORGANISM: Candida albicans
US-10-603-113-20275

```

Query Match	7.8%;	Score 123.5;	DB 6;	Length. 630;
Best Local Similarity	24.1%;	Pred. No. 0.25;		
Matches	77;	Conservative	51;	Mismatches 101; Indels 91; Gaps 18
QY	42	MKGDVITEHQF----	YEQVKS--	NPSAQVLLNMITQKFEKQ-----YGSGL 83
Db	189	LKG--ITERELAKSELETYKNSGLSPITSELAATKTV-	KSLKKKEKLEELQPLSGNKSREL 245	
QY	84	DD-----	KVDDITAEKKKQYGVYQAV-----	LSQAGMITETPKRAQITPSK 125
Db	246	EDYIQKSHSDISEKKALTDLDEKTKQFPDPSKKKLTLELNDLSTFKKELETETKQ--	TSK 303	
QY	126	LEVEL-----	AVKVAEAEI--	TDQAKKARDEYTPDYTAIIILNN--EDKAK---- 169
Db	304	FKNLEBKDKETIVAKNELLELLKDKNSGAKKELLEKVSKLSEITELISLELDDKKSVMKQ 363		
QY	170	--EVLKAKAGADPAQLAKDNST--	-DEKTEKENGELT--	-FDSASTEVPQVYKKA 218
Db	364	HDELKEQYKREKNGEOLQVTKDYSTTKKLKDELDELQKELDALSPFQDKETYSAKLTOSTS--	421	
QY	219	AFALDDVGVSVYITATGTQAVSSOYIVKLYTKK--	TEKSSNIDDYKREKIKTVIILQKONDS 277	
Db	422	----DLKANKKLLILISEKEKTEQOELEKTKHQAHELEKLID--	KNKDS 465	
QY	278	TEVOSIIKKELOAAINIKYKD 297		
Db	466	SKTNS----DLKANKKVLD 481		

RESULT 9
 US-10-299-636-96
 Sequence 96, Application US/10299636
 GENERAL INFORMATION:
 APPLICANT: Billes, David E
 APPLICANT: McDaniel, Larry S
 APPLICANT: Swiatlo, Edwin
 APPLICANT: Yoether, Janet
 APPLICANT: Crain, Marilyn J
 APPLICANT: Hollingshead, Susan
 APPLICANT: Tarr, Rebecca
 APPLICANT: Brooks-Walter, Alexis
 TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEINS AND USES THEREOF
 FILE REFERENCE: 57909/361
 CURRENT APPLICATION NUMBER: US/10/299,636
 CURRENT FILING DATE: 2002-11-19
 PRIOR APPLICATION NUMBER: 08/714,741
 PRIOR FILING DATE: 1996-09-16
 PRIOR APPLICATION NUMBER: 08/529,055
 PRIOR FILING DATE: 1995-09-15
 NUMBER OF SEQ ID NOS: 111
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 96

```
; LENGTH: 588
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-299-636-96
```

Query Match	7.7%;	Score 123;	DB 6;	Length 588;
Best Local Similarity	25.1%;	Pred. NO. 0.25;		
Matches 64;	Conservative 44;	Mismatches 103;	Indels 44;	Gaps 11

[illegible]

```

RESULT 10
US-10-603-113-14747
: Sequence 14747, Application US/10603113
: GENERAL INFORMATION:
: APPLICANT: Keith Weinstock et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
: TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.132
: CURRENT APPLICATION NUMBER: US/10/603,113
: CURRENT FILING DATE: 2003-06-24
: PRIOR APPLICATION NUMBER: US/09/248,796
: PRIOR FILING DATE: 1999-02-12
: NUMBER OF SEQ_ID NOS: 28206
SEQ_ID NO 14747
: LENGTH: 901
: TYPE: PRT
: ORGANISM: Candida albicans
US-10-603-113-14747

```

[illegible]

```

QY      270 LTQKQNDSTVFQSIIGKELQANIKVKDQAFQNFQYQIGG 310
          :|      | | | : :      |||  ||
Db      791 DDDDLDDLDISPEIRKHLQSPV-----YFTSLAGG 823

```

RESULT 11
US-10-273-573-9033
; Sequence 9033, Application US/10273573

```

1  APPLICANT: Hyseq, Inc
2  TITLE OF INVENTION: NOVEL MACROPHAGE NUCLEIC ACIDS AND POLYPEPTIDES
3  FILE REFERENCE: 21272-066
4  CURRENT APPLICATION NUMBER: US/10/273,573
5  CURRENT FILING DATE: 2002-10-18
6  PRIOR APPLICATION NUMBER: 09/522,929
7  PRIOR FILING DATE: 2000-04-18
8  PRIOR APPLICATION NUMBER: 09/770,160
9  PRIOR FILING DATE: 2001-01-26
10 NUMBER OF SEQ ID NOS: 10994
11
12 SOFTWARE: Custom
13
14 SEQ ID NO 9033
15
16 LENGTH: 802
17
18 TYPE: PRT
19
20 ORGANISM: Homo sapiens
21
22 FEATURE:
23
24 NAME/KEY: DOMAIN
25 LOCATION: (653)..(707)
26 OTHER INFORMATION: Kinesin light chain repeat proteins domain identified by
27 OTHER INFORMATION: EMATRIX, accession number BL01160B, p-value=5.424e-09, raw sco
28 OTHER INFORMATION: 19.54
29
30 US_10-273-573-9033

```

	Query Match	7.7%	Score 121.5	DB 6	length 802:
	Best Local Similarity	20.4%	Pred. No. 0.46:		
	Matches	60;	Conservative	66;	Mismatches 107; Indels 61; Gaps 11
QY	35 EGADLISMKGDVITEHGFVEQVKSNSPAQOVLNNTIQKVEEKQGSLELDKEVDVTIAE	94			
	: : : : : : :				
Db	431 ELGELIPLRQKVAQERTAQQLRAEKASYAQLSML-----KKAGGLAAE	475			
QY	95 EKKQY-----ENYRVLSQAQMLETTRKAQIR----TSKLEVELAV	131			
	: : : : : : :				
Db	476 ENRGIGERANLRQFLLEVLDQAREKEYQVELAAVRADETRRLAEVQREASTARELEVMT	535			
QY	132 KKVAAEILTDAAYKKAPDEYTPPDYAQTIIRLN--NEDAKEVLEKAKAGADFPAQLACDN	189			
	: : : : : : :				
Db	536 AKYGCAKVKYLEKRQRFQEROKKLTAQVEOLEFQREQTKQWELSK--KLADSDASKQV	594			
QY	190 STDEK-TKENGSELTFDSATFEVPBQVKKAFALDVDSVDITATGTGNVSSQYYIVYL	248			
	: : : : : : :				
Db	595 QQKLKAAVOAQGE-----SQQEARILDAQ-----INELQAOALSQEQAAEHYKLM	640			
QY	249 TK-KTE---KSSNIDDKYEKLKTLYLTQRONDSTFVS--TIKELQANIKRYD	297			
	: : : : : : :				
Db	641 EKATHTHDAAQQNDQLDLQJRTLEFDQLQKNKELRALBAERGRCHETLDQAGLKTRKE	694			

```

RESULT 12
US-10-273-9040
: Sequence 9040, Application US/10273573
: GENERAL INFORMATION:
: APPLICANT: Hyseq, Inc
: TITLE OF INVENTION: NOVEL MACROPHAGE NUCLEIC ACIDS AND POLYPEPTIDES
: FILE REFERENCE: 212/2-066
: CURRENT APPLICATION NUMBER: US/10/273,573
: CURRENT FILING DATE: 2002-10-18
: PRIOR APPLICATION NUMBER: 09/522,929
: PRIOR FILING DATE: 2000-04-18
: PRIOR APPLICATION NUMBER: 09/770,160
: PRIOR FILING DATE: 2001-01-26
: NUMBER OF SEQ ID NOS: 10994
: SOFTWARE: Custom
: SEQ ID NO 9040

```

```

; LENGTH: 1154
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (169)..(217)
; OTHER INFORMATION: Intermediate filaments proteins domain identified by eMatrix,
; OTHER INFORMATION: accession number BL00226b, p-value=3.093e-09, raw score of 23.86
; US-10-299-573-9040

```

Query Match 7.7%; Score 121.5; DB 6; Length 1154;
Best Local Similarity 20.4%; Pred. No. 0.73; Indels 61; Gaps 11;

Matches 60; Conservative 66; Mismatches 107; Indels 61; Gaps 11;

```

OY 35 EGADLISKGDVITEHOFYEQVSNPSAQVLLNMTIQKVEKQSGELDKEDVDITAE 94
DB 796 ELGLIPLRQKVAEOERTAOQLRAEKASYAEQLSML-----KKAHGLAE 840
OY 95 EKQYQ-----ENQRYLSQAGMTLETRKAQIR-----TSKLEVLAV 131
DB 841 ENRGLGERALNGRQFLEVLDAQREKIVQELAAVRADAETRLAEVQREASTARELEVMT 900
OY 132 KVAEAEITDEAVKKAPEDEYTPDVAQIIRLN--NEDKAKEVLEKKAEGADPAQLAKON 189
DB 901 AKYEGAKVYVLEERQRFQEROKLTQAQVEQLEFQHQTKQVLELSK-KLADSDQSKVO 959
OY 190 STDEK-TKENGGETPDSASTVEPEQVKKAFALDVGSDVITANGTOAYSSQYIVKL 248
DB 960 QQKLKAVQAQGE-----SQEQAORLQAQ-----INLEQAQLSQKEQAHEHYKLM 1005
OY 249 TK-KTE---KSSNIDYKEKLTQVITLQKONDSFVOS--TIGKELOANIKYVD 297
DB 1006 ERAKTHYDAKKQONQELQELRLDEQLQENKELRAERELGHELOAQGLTKTE 1059

```

RESULT 13

```

; Sequence 103, Application US/10299636
; GENERAL INFORMATION:
; APPLICANT: Briles, David E
; APPLICANT: McDaniel, Larry S
; APPLICANT: Swiatlo, Edwin
; APPLICANT: Yotner, Janet
; APPLICANT: Crain, Marilyn J
; APPLICANT: Hollingshead, Susan
; APPLICANT: Tarr, Rebecca
; APPLICANT: Brooks-Walter, Alexis
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEINS AND USES THEREOF
; FILE REFERENCE: 57909/361
; CURRENT APPLICATION NUMBER: US/10/299,636
; PRIOR FILING DATE: 2002-11-19
; PRIOR APPLICATION NUMBER: 08/714,741
; PRIOR FILING DATE: 1996-09-16
; PRIOR APPLICATION NUMBER: 08/529,055
; NUMBER OF SEQ ID NOS: 111
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 103
; LENGTH: 336
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; US-10-299-636-103

```

Query Match 7.5%; Score 119; DB 6; Length 336;
Best Local Similarity 21.9%; Pred. No. 0.24; Indels 88; Gaps 11;

Matches 70; Conservative 45; Mismatches 116; Indels 88; Gaps 11;

```

OY 4 YLKGVDMMKKTLAGITLISVAITLAAKSGEGADLISKGDVITEHOFYEQVSNPSAQ 63
DB 27 HTQVNLNLIK--SAIKTKYIVLELSVLEKNSKKEELSKTVAELT--AAEFQFKDKLKP 82
OY 64 QVLLNMTIQKVEKQSGELDKEDVDITAEKQYQENYGRVLSQAGMTLETRKAQIRT 123

```

```

DB 83 E-----KVAEAEKKVEEAKKANDQKEEDRNPTN-----TYKTLLELAE 125
OY 124 SKLVELAVKVAEAEITDEAVKKAPEDEYTPDVAQIIRLNEDKAKEVLEKKAEGADPA 183
DB 126 SDV-----KVAEAE--ELVKEANE-----SRNEKIKQAEEKESKKAET 166
OY 184 QLKONDSDEKTEKENGGETPDSASTVEPEQVKKAFALDVGSDVITANGTOAYSSQY 243
DB 167 RLKE-IKTDKKAE-----EAKKRAESEEKKAFAASKLDPTK----- 203
OY 244 YIVKLTKTEKSSNIDYKEKLTQVITLQKONDSFVOSIIGKELOANIKYKDAQONI 303
DB 204 --AKLSKLELSDPKIDE-----LDAETAKLEVLQKDAEGNN 238
OY 304 FQYIGGGSSSSSTISNE 322
DB 239 VEAYFEKLEKTTAEKKA 257

```

RESULT 14

```

; Sequence 105, Application US/10299636
; GENERAL INFORMATION:
; APPLICANT: Briles, David E
; APPLICANT: McDaniel, Larry S
; APPLICANT: Swiatlo, Edwin
; APPLICANT: Yotner, Janet
; APPLICANT: Crain, Marilyn J
; APPLICANT: Hollingshead, Susan
; APPLICANT: Tarr, Rebecca
; APPLICANT: Brooks-Walter, Alexis
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEINS AND USES THEREOF
; FILE REFERENCE: 57909/361
; CURRENT APPLICATION NUMBER: US/10/299,636
; PRIOR FILING DATE: 2002-11-19
; PRIOR APPLICATION NUMBER: 08/714,741
; PRIOR FILING DATE: 1996-09-16
; PRIOR APPLICATION NUMBER: 08/529,055
; NUMBER OF SEQ ID NOS: 111
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 105
; LENGTH: 354
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; US-10-299-636-105

```

Query Match 7.5%; Score 119; DB 6; Length 354;
Best Local Similarity 25.8%; Pred. No. 0.25; Indels 32; Gaps 10;

Matches 59; Conservative 43; Mismatches 95; Indels 32; Gaps 10;

```

OY 86 KEVDITAEKQYQENYGRVLSQAGMTLETRKAQIRTSKLVELAVKVAEAEITDEAYK 145
DB 17 KALDDAKAAQK-YDEQOKTEEKAAL-----EKA--ASEEMKAAVAQAAYL--AVQ 65
OY 146 KAFDEYTPDVAQIIRLNEDKAKEVLEKAAE-----GADFAQLAKNDSDEKTEK 197
DB 66 QATDRAKADADAKKI-----DEAKKREEFAKTFNFVAVWVPEPDLAETKKKSEANQ 120
OY 198 NGGEIT--FDSASTVEPEQVKKAFALDVGSDVITATQVASSQYIVKLTKTEKS 255
DB 121 KAPELITKKEAEKAKKEAEKKAETAKQKVDAAEVAPQAKIAELQVHRLQELKEIDE 180
OY 256 SNIDY-KERLKTQV---LTQKONDSFVOSIIGK--ELOANIKYKDD 298
DB 181 SESEDYAKEGFRAPLOSILDAKAKAKLSKLELSDKIDELDAETAKLEDO 229

```

RESULT 15

```

; Sequence 3, Application US/10257377
; GENERAL INFORMATION:
; APPLICANT: B. Michael R. Woloski,

```

APPLICANT: Ashley M. Williams,
APPLICANT: Terrance J. Sereda,
APPLICANT: Deanna J. Wiebe
TITLE OF INVENTION: RHAM Peptide Conjugates
FILE REFERENCE: BERE-100105
CURRENT APPLICATION NUMBER: US/10/257,377
CURRENT FILING DATE: 2001-04-20
PRIOR APPLICATION NUMBER: US 60/198,613
PRIOR FILING DATE: 2000-04-20
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3
LENGTH: 713
TYPE: PRT
ORGANISM: Rattus norvegicus
US-10-257-377-3

Query Match 7.5%; Score 119; DB 6; Length 713;
Best Local Similarity 20.4%; Pred. No. 0.6;
Matches 78; Conservative 69; Mismatches 126; Indels 110; Gaps 16;

QY 31 SKGSEADLISAKGVIT-----EHQFYQVKSNPQAQVLL----- 67
DB 305 SKDREAEESLSAEMOYLTEKLLERQVEYELQONELOSLLQOEKLSAHLQOQLCSFQ 364
QY 68 -NNTIOK-VFEKOYGSGLDDEVDPTIAEKKOYGENYORVLSQAGMTLE--TRKAQIR 123
DB 365 EEMTSERNYFKQELALDEL---DAVQOKKEQSEKIVKOLEEETKSTABQLRLDDLK 421
QY 124 SKLVELAVKVAEELTDEAYKRAF--EYTPDVTAQI-----IRLNN---- 164
DB 422 EKEIELEKRTAAHAQATVIAQEKYSDTAQTLRDVTAQLESYKSTLKEIEDLKENLTLQ 481
QY 165 -----EDKAKVLEKAKAEGADFAQAKD--NS-----TP- 192
DB 482 EKVAAAEKREVDVQOQIL--TAESTNQEVAKVYODLQNSSTLKEAIKEITSSYLEKITDL 540
QY 193 -----EKTRENGCEIT-FDSASTVEPEQVYK-----AFALDVGV 227
DB 541 ONQLRQONEDFRKQLEEGAKMTEKETAATVETLMEINKRRLLYEELFDKTKPFQOQDDAF 600
QY 228 SDVTATGTQAYSSQYIVKLTAKTEKSSNIDDKYKELKTVILTKQNDSTFVQSIIGK- 286
DB 601 EAEKQALLNEHGATQOLSKIRDSYAQLGHQNLKQIKHVYKLDKDNSQ--LKSEVSKL 658
QY 287 ELQANIKYKDAQFQNIPTQYIG 309
DB 659 RSQLAARKQNELRLQGEIDKALG 681

Search completed: August 29, 2003, 10:44:31
Job time : 26 secs

BEST AVAILABLE COPY

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2003, 10:19:19 ; Search time 83 Seconds
(Without alignments)
615.783 Million cell updates/sec

Title: US-10-049-473a-2
Perfect score: 1588
Sequence: 1 SNTYLGVDKMKRLAGAIR.....IFQYIGGDSSTSSSTSSNE 322

Scoring table: BLAST62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03 :
1: /SIDSL/gcgdata/geneseq/emb1/AA1980.DAT :
2: /SIDSL/gcgdata/geneseq/emb1/AA1981.DAT :
3: /SIDSL/gcgdata/geneseq/emb1/AA1982.DAT :
4: /SIDSL/gcgdata/geneseq/emb1/AA1983.DAT :
5: /SIDSL/gcgdata/geneseq/emb1/AA1984.DAT :
6: /SIDSL/gcgdata/geneseq/emb1/AA1985.DAT :
7: /SIDSL/gcgdata/geneseq/emb1/AA1986.DAT :
8: /SIDSL/gcgdata/geneseq/emb1/AA1987.DAT :
9: /SIDSL/gcgdata/geneseq/emb1/AA1988.DAT :
10: /SIDSL/gcgdata/geneseq/emb1/AA1989.DAT :
11: /SIDSL/gcgdata/geneseq/emb1/AA1990.DAT :
12: /SIDSL/gcgdata/geneseq/emb1/AA1991.DAT :
13: /SIDSL/gcgdata/geneseq/emb1/AA1992.DAT :
14: /SIDSL/gcgdata/geneseq/emb1/AA1993.DAT :
15: /SIDSL/gcgdata/geneseq/emb1/AA1994.DAT :
16: /SIDSL/gcgdata/geneseq/emb1/AA1995.DAT :
17: /SIDSL/gcgdata/geneseq/emb1/AA1996.DAT :
18: /SIDSL/gcgdata/geneseq/emb1/AA1997.DAT :
19: /SIDSL/gcgdata/geneseq/emb1/AA1998.DAT :
20: /SIDSL/gcgdata/geneseq/emb1/AA1999.DAT :
21: /SIDSL/gcgdata/geneseq/emb1/AA2000.DAT :
22: /SIDSL/gcgdata/geneseq/emb1/AA2001.DAT :
23: /SIDSL/gcgdata/geneseq/emb1/AA2002.DAT :
24: /SIDSL/gcgdata/geneseq/emb1/AA2003.DAT :
25: /SIDSL/gcgdata/geneseq/emb1/AA2004.DAT :
26: /SIDSL/gcgdata/geneseq/emb1/AA2005.DAT :
27: /SIDSL/gcgdata/geneseq/emb1/AA2006.DAT :
28: /SIDSL/gcgdata/geneseq/emb1/AA2007.DAT :
29: /SIDSL/gcgdata/geneseq/emb1/AA2008.DAT :
30: /SIDSL/gcgdata/geneseq/emb1/AA2009.DAT :
31: /SIDSL/gcgdata/geneseq/emb1/AA2010.DAT :
32: /SIDSL/gcgdata/geneseq/emb1/AA2011.DAT :
33: /SIDSL/gcgdata/geneseq/emb1/AA2012.DAT :
34: /SIDSL/gcgdata/geneseq/emb1/AA2013.DAT :
35: /SIDSL/gcgdata/geneseq/emb1/AA2014.DAT :
36: /SIDSL/gcgdata/geneseq/emb1/AA2015.DAT :
37: /SIDSL/gcgdata/geneseq/emb1/AA2016.DAT :
38: /SIDSL/gcgdata/geneseq/emb1/AA2017.DAT :
39: /SIDSL/gcgdata/geneseq/emb1/AA2018.DAT :
40: /SIDSL/gcgdata/geneseq/emb1/AA2019.DAT :
41: /SIDSL/gcgdata/geneseq/emb1/AA2020.DAT :
42: /SIDSL/gcgdata/geneseq/emb1/AA2021.DAT :
43: /SIDSL/gcgdata/geneseq/emb1/AA2022.DAT :
44: /SIDSL/gcgdata/geneseq/emb1/AA2023.DAT :
45: /SIDSL/gcgdata/geneseq/emb1/AA2024.DAT :

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1588	100.0	322	22	AA873245
2	1541	97.0	313	24	ABU01392
3	1541	97.0	314	21	AAV81675
4	916	57.7	213	19	AAW5079
5	916	57.7	213	23	ABP54573
6	736	46.3	309	23	ABP50803
7	736	46.3	309	23	ABP30888
8	717	45.2	298	23	ABP3081
9	696	43.8	351	23	ABP28729

10	625	39.4	317	23	ABP30009	Streptococcus poly
11	605.5	38.1	342	20	AAV00100	Enterococcus faeca
12	605.5	38.1	342	20	ABP43319	E faecalis EF049 p
13	605.5	38.1	342	24	ABU13598	Enterococcus faeca
14	562.5	35.4	322	24	AAV00101	Enterococcus faeca
15	562.5	35.4	322	23	ABP43320	E faecalis EF049 a
16	562.5	35.4	322	24	ABU13599	Enterococcus faeca
17	516	32.5	217	23	ABP28728	Streptococcus poly
18	511	32.2	308	23	ABW50074	Lactococcus lactis
19	393.5	24.8	293	23	ABW94437	Listeria monocytog
20	375.5	23.6	299	22	AAV3246	Protease maturatio
21	374	23.6	299	22	AAV3248	Protease maturatio
22	328	20.7	294	23	ABP47961	Listeria monocytog
23	322	20.3	74	19	AAV38711	Streptococcus pneu
24	322	20.3	74	19	AAV38711	Streptococcus pneu
25	212.5	13.4	320	24	ABU18994	Pathogen specific
26	210	13.2	330	23	ABP38966	Staphylococcus epl
27	204	12.8	320	20	AAW89744	Staphylococcus aur
28	173	10.9	42	18	AAV11343	S. pneumoniae prot
29	155.5	9.8	336	20	AAV19969	B. burgdorferi ant
30	153.5	9.7	317	20	AAV19969	B. burgdorferi ant
31	136.5	8.6	684	16	AAV3912	Streptococcus pneu
32	135.5	8.5	619	15	AAV3437	Pneumococcal surfa
33	135.5	8.5	619	17	AAV87598	Pneumococcal surfa
34	135.5	8.5	619	20	AAV1838	Streptococcus pneu
35	135.5	8.5	619	23	AAE18782	S. pneumoniae Rxi
36	135.5	8.5	648	19	AAV70336	Pneumococcal surfa
37	135.5	8.5	648	19	AAV62274	Streptococcus pneu
38	135.5	8.5	648	20	AAV1837	Streptococcus pneu
39	135.5	8.5	648	20	AAV87879	A pneumoniae trun
40	135.5	8.5	653	17	AAW92456	S. pneumoniae trun
41	130.5	8.2	619	17	AAV86911	Pneumococcal surfa
42	130.5	8.2	2478	22	AAU34370	Staphylococcus aur
43	130.5	8.2	2478	22	AAU34370	Staphylococcus aur
44	130.5	8.2	2478	24	ABU19002	Pathogen specific
45	130	8.2	288	14	AAV35436	B. burgdorferi str

ALIGNMENTS

RESULT 1	AA873245	standard; Protein: 322 AA.
ID	AA873245;	
XX	AC	AA873245;
XX	DT	14-MAY-2001 (first entry)
XX	DE	Protease maturation protein, Pmp.
XX	KW	Antibacterial; protease maturation protein; Pmp; bacterial infection;
XX	KW	vaccine.
XX	OS	Streptococcus pneumoniae.
XX	PN	EP1075841-AA
XX	PD	14-FEB-2001.
XX	XX	13-AUG-1999; 99EP-0202640.
XX	PR	13-AUG-1999; 99EP-0202640.
XX	PA	(UYRO-) UNIV ROTTERDAM ERASMUS.
XX	PI	De Groot R, Hermans PMM;
XX	DR	WPI: 2001-236756/25.
XX	DR	N-PSDB: AAF75748.
XX	PT	New vaccine comprising a protease maturation protein from Streptococcus pneumoniae useful for treating or preventing microbial infections,

PT specifically streptococcal infections

XX Claim 4; Fig 1; 18pp; English.

PS Streptococcus pneumoniae is a major cause of invasive diseases such as

CC meningitis, bacteraemia and pneumonia, as well as non-invasive diseases

CC such as acute otitis media and sinusitis. The present invention relates

CC to a vaccine comprising Protease Maturation Protein (Pmp) of

CC S. pneumoniae (the present sequence). The vaccine can be used for the

CC treatment of S. pneumoniae infections. Pmp is a surface-associated

CC protein.

CC

XX Sequence 322 AA;

SO

Query Match 100.0%; Score 1588; DB 22; Length 322;

Best Local Similarity 100.0%; Pred. No. 2,4e-113;

Matches 322; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SNTVLKGVDMKKKLLAGATITLLSVATLAACSGSGADLISMKGDVITEHQFYEQVKNP 60

DB 1 SNTVLKGVDMKKKLLAGATITLLSVATLAACSGSGADLISMKGDVITEHQFYEQVKNP 60

OY 61 SAQGVLLNMTIQKVEKQYSGELDDKEVDPTIAEKKQYGENYQVLSQAGMTLETRKAQ 120

DB 61 SAQGVLLNMTIQKVEKQYSGELDDKEVDPTIAEKKQYGENYQVLSQAGMTLETRKAQ 120

OY 121 IRRSKLVEIAVKKAAEELTDEAKKAFDETPVTAQIIRLNNEDEKAKVELEKAKEGA 180

DB 121 IRRSKLVEIAVKKAAEELTDEAKKAFDETPVTAQIIRLNNEDEKAKVELEKAKEGA 180

OY 181 DPAQLANDNSTDEKTEKNGEITFDSASTVEPEQVKAAPALDVGSDVITATGTQAYS 240

DB 181 DPAQLANDNSTDEKTEKNGEITFDSASTVEPEQVKAAPALDVGSDVITATGTQAYS 240

OY 241 SOYIYVLTITKTEKSSNIDYKREKLTIVILTKQONDSFVOSIIGKELQANIKVKDQAF 300

DB 241 SOYIYVLTITKTEKSSNIDYKREKLTIVILTKQONDSFVOSIIGKELQANIKVKDQAF 300

OY 301 QNIFTQYIIGGDSSTSTNE 322

DB 301 QNIFTQYIIGGDSSTSTNE 322

RESULT 2

ABU01392

ID ABU01392 standard; Protein; 313 AA.

XX

AC ABU01392;

DT 11-FEB-2003 (first entry)

XX

DE 5. pneumoniae type 4 strain protein from coding region #967.

XX

KM Bacterial meningitis; pneumonia; sepsis; otitis media;

KM ear infection; antiinflammatory; antibacterial; immunostimulant;

KM auditory; respiratory; gene therapy; vaccine.

XX

OS Streptococcus pneumoniae type 4 strain.

XX

PN MO200277021-A2.

XX

PD 03-OCT-2002.

XX

PF 27-MAR-2002; 2002WO-1B02163.

XX

PR 27-MAR-2001; 2001GB-0007658.

XX

PA (CHIR-) CHIRON SPA.

XX (GENO-) INST GENOMIC RES.

PI Masignani V, Tettelin H, Fraser C;

DR MPI; 2003-040579/03.

DR N-PSDB; ABX06679.

XX

PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,

PT useful as medicaments for treating or preventing a disease or infection

PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media

PT or ear infection

PS

XX Claim 1; SEQ ID No 1934; 56pp; English.

CC The invention relates to a protein comprising or having at least 50%

CC identity to any of the 2469 amino acid sequences, identified in the

CC specification (available on a computer readable format), or its fragment,

CC expressed from 2469 of 2489 identified DNA coding regions from the

CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as

CC AB856454. Also included are an antibody which binds one of the

CC proteins, treating a patient by administering the protein, DNA or

CC antibody (in a composition), a kit comprising first and second primers,

CC which are the nucleic acid cited above or fragments between nucleotides

CC 8-100 of a sequence not defined in the specification, for amplifying a

CC target sequence contained within a Streptococcus nucleic acid sequence,

CC where the first primer is substantially complementary to the target

CC sequence and the second primer is substantially complementary to the

CC complement of the target sequence, and where the parts of the primers

CC having substantial complementarity define the termini of the target

CC sequence to be amplified, assay comprising contacting a test compound

CC with the protein, and determining whether the test compound binds to the

CC protein and a Streptococcus pneumoniae bacterium, where one or more

CC genes encoding the proteins has been rendered inactive. The proteins,

CC nucleic acid molecules, antibody and compositions are useful as

CC medicaments for treating or preventing a disease or infection due to

CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,

CC sepsis, otitis media or ear infection. They are also useful in developing

CC vaccines, diagnostics and antibiotics. The methods are useful for

CC identifying immunodominant proteins. The present sequence is one of

CC the 2469 proteins expressed by the identified coding regions from the

CC genomic sequence.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences.

XX

SO Sequence 313 AA;

Query Match 97.0%; Score 1541; DB 24; Length 313;

Best Local Similarity 100.0%; Pred. No. 9e-110;

Matches 313; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 MKKLLAGATITLLSVATLAACSGSGADLISMKGDVITEHQFYEQVKNPSPAQVLLNM 69

DB 1 MKKLLAGATITLLSVATLAACSGSGADLISMKGDVITEHQFYEQVKNPSPAQVLLNM 69

OY 70 TIQKVEKQYSGELDDKEVDPTIAEKKQYGENYQVLSQAGMTLETRKAQIRTSKIVEL 129

DB 70 TIQKVEKQYSGELDDKEVDPTIAEKKQYGENYQVLSQAGMTLETRKAQIRTSKIVEL 129

OY 61 TIQKVEKQYSGELDDKEVDPTIAEKKQYGENYQVLSQAGMTLETRKAQIRTSKIVEL 120

DB 61 TIQKVEKQYSGELDDKEVDPTIAEKKQYGENYQVLSQAGMTLETRKAQIRTSKIVEL 120

OY 130 AVKKAFAELTDEAKKAFDETPVTAQIIRLNNEDEKAKVELEKAKGAPQALANDN 189

DB 130 AVKKAFAELTDEAKKAFDETPVTAQIIRLNNEDEKAKVELEKAKGAPQALANDN 189

OY 121 AVKKAFAELTDEAKKAFDETPVTAQIIRLNNEDEKAKVELEKAKGAPQALANDN 180

DB 121 AVKKAFAELTDEAKKAFDETPVTAQIIRLNNEDEKAKVELEKAKGAPQALANDN 180

OY 190 STDEKTEKNGEITFDSASTVEPEQVKAAPALDVGSDVITATGTQAYSQYIYVLT 249

DB 190 STDEKTEKNGEITFDSASTVEPEQVKAAPALDVGSDVITATGTQAYSQYIYVLT 249

OY 181 STDEKTEKNGEITFDSASTVEPEQVKAAPALDVGSDVITATGTQAYSQYIYVLT 240

DB 181 STDEKTEKNGEITFDSASTVEPEQVKAAPALDVGSDVITATGTQAYSQYIYVLT 240

OY 250 KTEKSSNIDYKREKLTIVILTKQONDSFVOSIIGKELQANIKVKDQAFONIFTQYIG 309

DB 250 KTEKSSNIDYKREKLTIVILTKQONDSFVOSIIGKELQANIKVKDQAFONIFTQYIG 309

OY 241 KTEKSSNIDYKREKLTIVILTKQONDSFVOSIIGKELQANIKVKDQAFONIFTQYIG 300

DB 241 KTEKSSNIDYKREKLTIVILTKQONDSFVOSIIGKELQANIKVKDQAFONIFTQYIG 300

OY 310 GGDSSSSSTSTNE 322

DB 310 GGDSSSSSTSTNE 322

RESULT 3

Db			241 KRTKSSNIDYKKKLTVILITOKNDSTFVQSIIGKELOAAANKVAQDAFONIFTYIG	300
QY			310 GGDSSSSSSSTSNE	322
Db			301 GGDSSSSSSTSNE	313
RESULT 4				
AAM5079	ID		AAM5079 standard; Protein: 213 AA.	
XX				
AC			AAM5079;	
XX				
DT			02-OCT-1998 (first entry)	
XX				
DE			Streptococcus pneumoniae SP0021 protein.	
XX				
KM			Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;	
RW			detection; pneumonia; otitis media; meningitis.	
XX				
OS			Streptococcus pneumoniae.	
PX				
PN			M09818930-A2.	
PD			07-MAY-1998.	
PF			30-OCT-1997; 97WO-US19422.	
PR			31-OCT-1996; 96US-0029960.	
FA			(HUMA-) HUMAN GENOME SCI INC.	
XX				
PI			Choi GH, Hromockyj A, Johnson LS, Kunsch CA;	
DR			WPI: 1998-272224/24.	
XX			N-PDB: AAV27340.	
PT			Nucleic acid encoding antigenic peptide(s) from Streptococcus	
PT			pneumoniae - or their epitope-containing fragments, useful in	
PT			protective or therapeutic vaccines, and for diagnosis	
PS			Claim 11; Page 55; 118pp; English.	
XX				
CC			The present sequence represents a protein from Streptococcus pneumoniae.	
CC			The nucleic acid sequence encoding the Streptococcus pneumoniae protein	
CC			can be useful in vaccines for inducing protective antibodies against	
CC			Streptococcus pneumoniae, for treatment or prevention of infection e.g,	
CC			pneumonia, otitis media or meningitis. Probes based on the nucleic acid	
CC			are used to detect Streptococcus infection (by usual hybridisation or	
CC			amplification methods), also for isolating Streptococcus genes or their	
CC			allelic variants. The protein can be used similarly to detect specific	
CC			antibodies in standard immunoassays, especially for diagnosing or	
CC			monitoring infections. Antibodies which bind the protein are used to	
CC			detect corresponding antigens, to purify the protein and for passive	
CC			immunisation (optionally coupled to a toxin). Vaccines are administered,	
CC			e.g. by injection, orally or through the skin, typically at 0.01-1000	
CC			(especially 10-300) mu g/ml per dose.	
XX				
SQ			Sequence 213 AA:	
Query Match		57.7%;	Score 916; DB 19; Length 213;	
Best Local Similarity		93.6%;	Pred. No. 3,3e-62;	
Matches 190; Conservative		0; Mismatches	5; Indels	8; Gaps
2;				
QY			31 SKSGEGADLIMKGDVITEHFQFYQVSNPBAOYLNMTOAKFEKYGSELDKREVD	90
Db			1 SKSGEGADLIMKGDVITEHFQFYQVSNPBAOYLNMTOAKFEKYGSELDKREVD	60
QY			91 TIAEKKQYGENYRVLSQAQMTELETRKAQIRTSKIVELAVKVAEAELDEAKKAFDE	150
Db			61 TIAEKKQYGENYRVLSQAQMTELETRKAQIRTSKIVELAVKVAEAELDEAKKAFDE	120

```

OY 151 YTPDVTAGIIRLNNEDEKAEVLEKAKAGADFAOLAKDNSTDEKTEKNGEITFDASASTE 210
DB 121 YTPDVTAGIIRLNNEDEKAEVLEKAKAGADFAOLAKDNSTDEKTEKNGEITFDASASTE 180
OY 211 VP-EQVKKAAFA-----LDVD 225
DB 181 VPGASPKKPLFAFRGGMVFLDVD 203

RESULT 5
ABP54573
ID ABP54573 standard; Protein; 213 AA.
XX
AC ABP54573;
XX
DT 04-SEP-2002 (first entry)
XX
DE S. pneumoniae SP021 protein sequence SEQ ID NO:34.
XX
KM Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
XX antibacterial; Streptococcal infection; detection.
XX
OS Streptococcus pneumoniae.
XX
PN US2002061545-A1.
XX
PD 23-MAY-2002.
XX
PF 22-JAN-2001; 2001US-0765272.
XX
PR 30-OCT-1997; 97US-0961083.
XX
PA (CHOI/) CHOI G H.
PA (KUNS/) KUNSCH C A.
PA (BARA/) BARASH S C.
PA (DILL/) DILLON P J.
PA (DOUG/) DOUGHERTY B.
PA (FANN/) FANNON M R.
PA (ROSE/) ROSEN C A.
XX
PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
PI Rosen CA;
XX
DR WPI: 2002-479261/51.
DR N-PSDB: AB084808.
XX
PT New Streptococcus pneumoniae antigens, useful for detecting
PT Streptococcus and for preventing or attenuating disease caused by
PT Streptococcus infection.
XX
PS Claim 11; Page 24; 70pp; English.
XX
XX
AB084792 to AB084904 represents nucleic acids which encode the
CC Streptococcus pneumoniae antigens given in ABP54573 to ABP54669.
CC The S. pneumoniae antigens have antibacterial activity and can be
CC used in vaccines. The S. pneumoniae antigens can also be used to
CC prevent or attenuate a Streptococcal infection in an animal. The
CC polynucleotides encoding the S. pneumoniae antigens can be used to
CC detect Streptococcus nucleic acids. AB084905 to AB085130 represent
CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)
CC which are used in an example from the present invention.
XX
XX
SQ Sequence 213 AA;
XX
Query Match 57.7%; Score 916; DB 23; Length 213;
Best Local Similarity 93.6%; Pred. No. 3,3e-62;
Matches 190; Conservative 0; Mismatches 5; Indels 8; Gaps 2;
OY 31 SKGSEGADLISMKGDIYTHOFYEQYKSNPSAQOVLNNTIOKVFPEKOGSELDKXVD 90
DB 1 SKGSEGADLISMKGDIYTHOFYEQYKSNPSAQOVLNNTIOKVFPEKOGSELDKXVD 60
OY 91 TIAEKKQYGENYORVLSQAGMTLETTRKAQIRTSKILVELAVKVAEALIDEAYKKAPE 150

```

```

DB 61 TIAEKKQYGENYORVLSQAGMTLETTRKAQIRTSKILVELAVKVAEALIDEAYKKAPE 120
OY 151 YTPDVTAGIIRLNNEDEKAEVLEKAKAGADFAOLAKDNSTDEKTEKNGEITFDASASTE 210
DB 121 YTPDVTAGIIRLNNEDEKAEVLEKAKAGADFAOLAKDNSTDEKTEKNGEITFDASASTE 180
OY 211 VP-EQVKKAAFA-----LDVD 225
DB 181 VPGASPKKPLFAFRGGMVFLDVD 203

RESULT 6
ABP30803
ID ABP30803 standard; Protein; 309 AA.
XX
AC ABP30803;
XX
DT 02-JUL-2002 (first entry)
XX
DE Streptococcus polypeptide SEQ ID NO 10782.
XX
KM Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX group A streptococcus; Streptococcus pyogenes; antibacterial;
XX antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
OS Streptococcus agalactiae.
XX
PN W0200234771-A2.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-GB04789.
XX
PR 27-OCT-2000; 2000GB-0026333.
PR 24-NOV-2000; 2000GB-0028727.
PR 07-MAR-2001; 2001GB-0005640.
XX
XX
PA (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
PI Telford J, Maignani V, Margarit Ros YI, Grandi G, Fraser C;
PI Tettelein H;
XX
DR WPI: 2002-352536/38.
DR N-PSDB: ABN71434.
XX
PT New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein.
XX
XX
Claim 1; Page 4184; 4525pp; English.
XX
XX
The invention relates to a protein (ABP25413-ABP30895) from group B
CC Streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN6004-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
XX
XX
SQ Sequence 309 AA;

```


PA (GENO-) INST GENOMIC RES.
 XX Telford J, Maignani V, Margarit Ros YI, Grandi G, Fraser C;
 PI Tettein H;
 XX WPI: 2002-352536/38.
 DR N-PSDB; ABN70712.
 XX
 PT New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -
 XX
 PS Claim 1: Page 4058; 4535pp; English.
 XX
 CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC Streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (1), nucleic acids encoding (1), ABN6044-ABN71526 and
 CC antibodies that bind (1) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (1) are used to detect Streptococcus in a
 CC biological sample. (1) is used to determine whether a compound binds to
 CC (1). A composition comprising (1) or a nucleic acid encoding (1), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (1) may be used to recombinantly produce (1) and may be
 CC used in gene therapy. Antibodies to (1) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 CC
 XX Sequence 298 AA:
 SQ
 Query Match 45.2%; Score 717; DB 23; Length 298;
 Best Local Similarity 48.3%; Pred. No. 8.3e-47;
 Matches 145; Conservative 60; Mismatches 87; Indels 8; Gaps 3;
 OY 19 IITLIVATLAACS-KGSEGADLISMKDVTTEHOYFOVKSNSAQOVLNMTIQYFEK 77
 DB 1 MTLASVATLAACSGTNGNTVNTMKDITVSDYQVTKSAQOQSMITLISRVFDT 60
 OY 78 QYGESELDKDEVDITIAEKKOYGENYRVLSQAGMTLETAKQIRTSKVELAVKVAEA 137
 DB 61 QYGDVSKDKVSEAYNNKTAKGNSFSSALSGAGLTPEGRKOIRTMILVEYAVKEAAK 120
 OY 138 ELTDEAKKAFDETPVTAQIIRLNNEDEKAEVLEKAKAGADPAQLAND--NSIDEKT 195
 DB 121 ELTEPANKKEAYKNYTPETSVQVTKLDAEDKAKSVLKDKVAKADGADFAKIAKEKTTATDKV 180
 OY 196 KENGGEITFDASTEVEQVKKAAAFALDVGVSDVITATGTQAYSSQYIVYKLTKEKS 255
 DB 181 -----EKKFDSAGTTLPEKVEKMAAFKLDKNGVSDVSTYTKSYITIKYTDTEKK 235
 OY 256 SNIDYKREKLTIVILKQNDSTFVOSIIGKELQANIKVQDAFONIFQYIGGSDSS 315
 DB 236 SDMKSYKNRLKEVILKDKTSDRAFONKVISKALEKANVKIKDKAFAGIISQYATTS 295

OS Streptococcus pyogenes.
 XX W0200234771-A2.
 XX
 XX 02-MAY-2002.
 PD
 XX
 XX 29-OCT-2001; 2001MO-GB04789.
 PE
 XX 27-OCT-2000; 2000GB-0026333.
 PR 24-NOV-2000; 2000GB-0028727.
 PR 07-MAR-2001; 2001GB-0005640.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Telford J, Maignani V, Margarit Ros YI, Grandi G, Fraser C;
 PI Tettein H;
 XX WPI: 2002-352536/38.
 DR N-PSDB; ABN69360.
 XX
 PT New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -
 XX
 PS Claim 1: Page 3825; 4525pp; English.
 XX
 CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC Streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (1), nucleic acids encoding (1), ABN6044-ABN71526 and
 CC antibodies that bind (1) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (1) are used to detect Streptococcus in a
 CC biological sample. (1) is used to determine whether a compound binds to
 CC (1). A composition comprising (1) or a nucleic acid encoding (1), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (1) may be used to recombinantly produce (1) and may be
 CC used in gene therapy. Antibodies to (1) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 CC
 XX Sequence 351 AA:
 SQ
 Query Match 43.8%; Score 696; DB 23; Length 351;
 Best Local Similarity 44.7%; Pred. No. 4.2e-45;
 Matches 140; Conservative 62; Mismatches 105; Indels 6; Gaps 2;
 OY 13 KLAGAITLLSVATLAACSGSEGADLISMKDVTTEHOYFOVKSNSAQOVLNMTIQ 72
 DB 6 KLIASVATLAACSGTNDNTKRVISMKDITVSDYFNETKNTESOKAMINLVIS 65
 OY 73 KYFEKQYGESELDKDEVDITIAEKKOYGENYRVLSQAGMTLETAKQIRTSKVELAVK 132
 DB 66 RVFEAOYGDVSKDKVSEAYNNKTAKGNSFSSALSGAGLTPEGRKOIRTSKVELAVK 125
 OY 133 KYAEAEITDEAKKAFDETPVTAQIIRLNNEDEKAEVLEKAKAGADPAQLANDST 192
 DB 126 EAAKKELTTOEKKKAYESTPTMAVEMITLNDNEPFAKSVLEKAKAGADFAIAKEKTT- 184
 OY 193 EKTENGGEITFDASTEVEQVKKAAAFALDVGVSDVITATGTQAYSSQYIVYKLTKEKT 252
 DB 185 --TPKKYTYKFDSCATNPVTDVYKAASSLNCGISDVISYLDPTSYKKFYIVKTKKA 242
 OY 253 EKSSNIDYKREKLTIVILKQNDSTFVOSIIGKELQANIKVQDAFONIFQYIGGSD 312
 DB 243 EKSSWQEKKKLKAITIAESKDMNFONKYIANALDKRANVKIKDKAFAPANTIAANLQ 302
 OY 313 ---SSSSSTSN 322

us-10-049-473a-2.rag

Dd		68	LSLVIRSVETQYANKVSDSEKVEKAYKOQAQOIGSFYTAOGSLTPERTKKQIRITKL	127
Qy		127	VELAAVKVAEAE-LTDEAVKKAPEDEXYDPVTAAQIIRLNNEDEKAEVLEKAKAEGADFAQL	195
Dd		128	VEVAAVEAKKNENETISKDYRQAYDAVPTPMPAEIWMQFEKERDAAALAEVAKEGADFAAI	187
Qy		186	AKNSNF--DEKTENGELTEPDSASTEPEDQYKKAFAFLVDVGSDVDTATGTQAYSQY	243
Dd		188	AKERTTAAPAKKT---TYTFPSGETTLPAEVRAASGLKEGNSEIITLADPATSKRTY	242
Qy		244	YIYLKTRTKTESNIDYKEKLTAVILFQKONDSFTFYOSIIGKEIEJOANIKVKDOAFONI	303
Dd		243	HIIIVKTKKATKRADMAYOKRKLDIIVIGLKLPDPFNKYIAKALDXANVYIKDKAPANI	302
Qy		304	FTRY 307	
Dd		303	LQGF 306	
RESULT 11				
AA00100				
ID	AA00100	standard; Protein; 342 AA.		
XX	AA00100;			
AC	AA00100;			
XX				
DT	20-APR-1999	(first entry)		
XX				
DE	Enterococcus faecalis protein EF049.			
XX				
KM	Enterococcus faecalis; Infection; vaccine; immune response; diagnosis;			
RW	detection; attenuation; antigenic.			
XX				
OS	Enterococcus faecalis.			
PN	MO9850554-AA2.			
PD	12-NOV-1998.			
PF	04-MAY-1998;	98WO-US08959.		
PR	14-NOV-1997;	97US-0066009.		
ER	06-MAY-1997;	97US-0044031.		
XX	16-MAY-1997;	97US-0046655.		
PA	(HUMA-) HUMAN GENOME SCI INC.			
PI	Bailey C, Choi GH, Hromockyj A, Kunsch CA;			
DR	WPI; 1999-070095/06.			
NR	N-PSDB; AAX20090.			
PT	New isolated Enterococcus faecalis polynucleotides - used to develop			
XX	products for the detection of Enterococcus and for use in vaccines			
XX	for prevention or attenuation of Enterococcus infection			
PS	Claim 9; Page 126; 301pp; English.			
CC	The present sequence represents a protein isolated from			
CC	Enterococcus faecalis. The present invention describes genes, proteins			
CC	and antigenic polypeptides isolated from E. faecalis. The proteins can			
CC	be used in vaccines for preventing or attenuating an infection caused			
CC	by a member of the Enterococcus genus in an animal. They can also be			
CC	used for detecting Enterococcus antibodies in a sample. The nucleotide			
CC	sequences can be used for detecting Enterococcus nucleic acids.			
CC	Products from the present invention can also be used for screening			
CC	compounds to identify agonists and antagonists of E. faecalis protein			
CC	activity.			
SQ	Sequence . 342 AA;			
Query Match	38.1%; Score 605.5; DB 20; Length 342;			
Best Local Similarity	41.8%; Pstd No. 3,3e-38;			

CC bacterial genera. The polynucleotides are useful as probes for gene
CC mapping and for identifying E. faecalis in biological samples. Sequences
CC ABU13508-ABU13755 represent EF040 polypeptides of the invention.
CC Note: The sequence data for this patent can also be obtained from USPTO
CC at seqdata.uspto.gov/sequence.html.

XX Sequence 342 AA;

Query Match 38.1%; Score 605.5; DB 24; Length 342;
Best Local Similarity 41.8%; Pred. No. 3.3e-38;
Matches 132; Conservative 62; Mismatches 113; Indels 9; Gaps 5;

DB 10 MKKLLAGATLLSVATLAACSGSEADLIMKGDVITEHQFYEQVKNPSAQOVLNM 69
||||| : : : ||||| || : : ||||| || : : ||||| : : ||
1 MKKLLAAGAMAVFSLAACSSGSK--DIATMKGSTIVDDPFYNDIKQSTSQAFSQM 58
OY 70 TIOKVFQKQYSELDDEVDVDTIAEKKQ--YGENYQFVLSQAGMTLETRRAQIRTSKL 126
DB 59 VIKVFEEKYGKVDKXIOKNFDEAKQEVQAGGFSDALQAGLITKTFKQLKORAA 118
OY 127 VELAVKVAEELTDYAYKKAFDEYDPDTAQIIRLNEDKAKEVLEKAKAGADFAOLA 186
DB 119 YVAGLK--AHLKITDDELTAWASFPVEAQIIOVASEDDAKAV--KKEITGGDFTKTA 175
OY 187 KDNSTDEKTRKNGEITFDASTVEVEQYKKAFAFDVGVSDVITATGTOAYSSQYTY 246
DB 176 KKSSTDTATKKGCGKIKFDSQATVPAEYKKAFAFKKDEVESEPIATATMOTYTYV 235
OY 247 KLTKTEKSSNIDDKREKLTLYLTOKONDSTFVOSITKEILOANIKYKDOAFONIFQ 306
DB 236 KMTKNKAKNDMKPYEKEIKIAETKTLADQTFVSKVIDELKAANVKTIDDAFKNALAG 295
OY 307 YIGGGSSSSSSSTNS 322
DB 296 IM-QTESSASSSEKKE 310

RESULT 14
AAY00101

ID AAY00101 standard; Protein; 322 AA.

XX AAY00101;

DT 20-APR-1999 (first entry)

DE Enterococcus faecalis antigenic polypeptide fragment EF049.

KW Enterococcus faecalis; infection; vaccine; immune response; diagnosis;

KM detection; attenuation; antigenic.

OS Enterococcus faecalis.

PN WO9850554-A2.

PD 12-NOV-1998.

PF 04-MAY-1998; 98WO-US08959.

PR 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.

PR 16-MAY-1997; 97US-0046655.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Bailey C, Choi GH, Hromockyj A, Kunsch CA;

DR WPI; 1999-070095/06.

DR N-PSDB; AAX20091.

XX New isolated Enterococcus faecalis polynucleotides - used to develop
PT products for the detection of Enterococcus and for use in vaccines
PT for prevention or attenuation of Enterococcus infection

PS Claim 9; Page 127; 301pp; English.

XX The present sequence represents an antigenic polypeptide fragment
CC isolated from Enterococcus faecalis. The present invention describes
CC genes, proteins and antigenic polypeptides isolated from E. faecalis.
CC The proteins can be used in vaccines for preventing or attenuating an
CC infection caused by a member of the Enterococcus genus in an animal.
CC They can also be used for detecting Enterococcus antibodies in a sample.
CC The nucleotide sequences can be used for detecting Enterococcus nucleic
CC acids. Products from the present invention can also be used for
CC screening compounds to identify agonists and antagonists of E. faecalis
CC protein activity.

XX Sequence 322 AA;

Query Match 35.4%; Score 562.5; DB 20; Length 322;
Best Local Similarity 41.2%; Pred. No. 6e-35;
Matches 122; Conservative 58; Mismatches 107; Indels 9; Gaps 5;

DB 30 CSKSGEADLIMKGDVITEHQFYEQVKNPSAQOVLNMTIOKVFQKQYSELDDEVD 89
||| : : : ||||| || : : ||||| || : : ||||| : : ||
1 CSKSGK--DIATMKGSTIVDDPFYNDIKQSTSQAFSQMVIKVFEEKYGKVDKXIQ 58
OY 90 DTIAEKKQ--YGENYQFVLSQAGMTLETRRAQIRTSKLVELAVKVAEELTDYAYK 146
DB 59 KNEDEAKQEVQAGGFSDALQAGLITKTFKQLKORAAVDAGLK--AHLKITDDELKT 116
OY 147 AFDEYTPDTAQIIRLNEDKAKEVLEKAKAGADFAOLAKNSDEKTRKNGEITPDS 206
DB 117 AMASFPVEAQIIOVASEDDAKAV--KKEITGGDFTKIAKERSDTATKKGCGKIKFDS 175
OY 207 ASTEVEQYKKAFAFDVGVSDVITATGTOAYSSQYTYVKTLEKTEKSSNIDDKREKL 266
DB 176 QATVPAEYKKAFAFKKDEVESEPIATATMOTYTYVYVKKTKKAKNDMKPYEKEIK 235
OY 267 TYLTOKONDSTFVOSITKEILOANIKYKDOAFONIFQYIGGGSSSSSSSTNS 322
DB 236 KIAETKTLADQTFVSKVIDELKAANVKTIDDAFKNALAGV--QTESSASSSEKKE 290

RESULT 15
ABP43320

ID ABP43320 standard; Protein; 322 AA.

XX ABP43320;

DT 05-AUG-2002 (first entry)

DE E faecalis EF049 antigenic fragment.

KW Enterococcus; vaccine; gastrointestinal disease; diagnosis; antibiotic.

KM Enterococcus faecalis.

OS Enterococcus faecalis.

PN US2002045737-A1.

PD 18-APR-2002.

PF 04-MAY-1998; 98US-0071035.

PR 04-MAY-1998; 98US-0071035.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Choi GH, Bailey C, Hromockyj A, Kunsch CA;

DR WPI; 2002-425450/45.

DR N-PSDB; ABN98076.

XX New genes and polypeptides from Enterococcus faecalis, useful as
PT vaccines for preventing, treating or attenuating an infection caused by
PT a member of the Enterococcus genus in an animal, particularly E.
PT faecalis

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2003, 10:32:29 ; Search time 29 seconds
(without alignments)
469.796 Million cell updates/sec

Title: US-10-049-473a-2

Perfect score: 1588
Sequence: 1 SMYTKGVDMKKRLAGAIT.....IFQYIGGDSSTSSSTSN 322

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database:

Issued_Patents_AA:*
1: /cgn2_6/prodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/prodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	916	57.7	213	3	US-08-961-083-34
2	916	57.7	213	4	US-09-356-784-34
3	634	39.9	344	4	US-09-107-532A-6355
4	605.5	38.1	342	4	US-09-071-035-182
5	588.5	37.1	303	4	US-09-107-532A-6113
6	562.5	35.4	322	4	US-09-071-035-184
7	450.5	28.4	302	4	US-09-107-532A-4736
8	322	20.3	74	4	US-08-858-207A-508
9	210	13.2	330	4	US-09-134-001C-3811
10	147.5	9.3	447	4	US-09-252-991A-17422
11	137.5	8.7	443	4	US-09-328-352-7725
12	135.5	8.5	619	1	US-08-465-746-2
13	135.5	8.5	619	1	US-08-214-164-2
14	135.5	8.5	619	2	US-08-467-852A-3
15	135.5	8.5	619	2	US-08-246-636-2
16	135.5	8.5	619	2	US-08-247-491A-3
17	135.5	8.5	619	2	US-08-319-795-2
18	135.5	8.5	619	2	US-08-468-985-2
19	135.5	8.5	619	3	US-08-312-849-2
20	135.5	8.5	648	1	US-08-072-070-2
21	135.5	8.5	648	1	US-08-469-434-2
22	135.5	8.5	648	1	US-08-214-222-2
23	135.5	8.5	648	2	US-08-467-852A-2
24	135.5	8.5	648	2	US-08-468-718-2
25	135.5	8.5	648	2	US-08-247-491A-2
26	135.5	8.5	648	3	US-08-446-201-3
27	135.5	8.5	695	1	US-08-127-499A-23

28	135.5	8.5	695	1	US-08-482-847-23	Sequence 23, Appl
29	131.5	8.3	588	4	US-08-714-741-42	Sequence 42, Appl
30	131.5	8.3	1231	4	US-08-714-741-41	Sequence 41, Appl
31	129.5	8.2	864	4	US-08-714-741-40	Sequence 40, Appl
32	126.5	8.0	8991	4	US-08-714-741-32	Sequence 32, Appl
33	124	7.8	370	4	US-09-252-991A-29473	Sequence 29473, A
34	124	7.8	485	4	US-09-071-035-34	Sequence 34, Appl
35	123.5	7.8	274	3	US-08-479-017-8	Sequence 8, Appl
36	123.5	7.8	289	1	US-08-072-070-4	Sequence 4, Appl
37	123.5	7.8	289	1	US-08-469-434-4	Sequence 4, Appl
38	123.5	7.8	289	1	US-08-214-222-4	Sequence 4, Appl
39	123.5	7.8	289	2	US-08-467-852A-5	Sequence 5, Appl
40	123.5	7.8	289	2	US-08-468-718-4	Sequence 4, Appl
41	123.5	7.8	289	2	US-08-247-491A-5	Sequence 5, Appl
42	123.5	7.8	288	3	US-08-312-849-4	Sequence 4, Appl
43	123	7.7	288	3	US-08-446-201-4	Sequence 4, Appl
44	123	7.7	288	3	US-08-446-201-4	Sequence 4, Appl
45	123	7.7	288	4	US-09-308-375-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-961-083-34
Sequence 34, Application US/08961083
Patent No. 6159469

GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Aptamers and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Human Genome Sciences, Inc
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 213 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-34

Query Match 57.7%; Score 916; DB 3; Length 213;
Best Local Similarity 93.6%; Pred. No. 1e-67;
Matches 190; Conservative 0; Mismatches 5; Indels 8; Gaps 2;

31 SKSGEGADLISMKGVDVTEHOFYEQVKSNSAQOVLNMTIOVFEKQYSELDKREVD 90
|||||
1 SKSGEGADLISMKGVDVTEHOFYEQVKSNSAQOVLNMTIOVFEKQYSELDKREVD 60

QY 91 TTAEEKOYGENORVLSOAGMTLETRKAQIRTSKLVLAAYKVAEELTDEAYKKAFDE 150
 DB 61 TTAEEKOYGENORVLSOAGMTLETRKAQIRTSKLVLAAYKVAEELTDEAYKKAFDE 120
 QY 151 YTPDVTAQIIRLNNEDEKAEVLEKAKAEGADFAQIAKDNSTDEKTEKNGEITFDASASTE 210
 DB 121 YTPDVTAQIIRLNNEDEKAEVLEKAKAEGADFAQIAKDNSTDEKTEKNGEITFDASASTE 180
 QY 211 VP-EQYKKAFA-----LDVD 225
 DB 181 VPGASPKPLFAFRCGMVFLVD 203

RESULT 2
 US-09-536-784-34
 ; Sequence 34, Application US/09536784
 ; Patent No. 6534082
 ; GENERAL INFORMATION:
 ; APPLICANT: Choi et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
 ; NUMBER OF SEQUENCES: 452
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850

COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/536,784
 ; FILING DATE: 30-Oct-1997
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/961,083
 ; FILING DATE: OCT-30-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Michelle S. Marks
 ; REGISTRATION NUMBER: 41,971
 ; REFERENCE/DOCKET NUMBER: PB340P3
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 309-8512
 ; INFORMATION FOR SEQ ID NO: 34:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 213 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 34:
 ; US-09-536-784-34

Query Match 57.7%; Score 916; DB 4; Length 213;
 Best Local Similarity 93.6%; Pred. No. 1e-67;
 Matches 190; Conservative 0; Mismatches 5; Indels 8; Gaps 2;

QY 31 SKSEGDALISMKGDTTEHOFTFEQVKSNSAOQVLLNMTIOKVFEEQYSELDKDVDD 90
 DB 1 SKSEGDALISMKGDTTEHOFTFEQVKSNSAOQVLLNMTIOKVFEEQYSELDKDVDD 60
 QY 91 TTAEEKOYGENORVLSOAGMTLETRKAQIRTSKLVLAAYKVAEELTDEAYKKAFDE 150
 DB 61 TTAEEKOYGENORVLSOAGMTLETRKAQIRTSKLVLAAYKVAEELTDEAYKKAFDE 120
 QY 151 YTPDVTAQIIRLNNEDEKAEVLEKAKAEGADFAQIAKDNSTDEKTEKNGEITFDASASTE 210
 DB 121 YTPDVTAQIIRLNNEDEKAEVLEKAKAEGADFAQIAKDNSTDEKTEKNGEITFDASASTE 180

QY 211 VP-EQYKKAFA-----LDVD 225
 DB 181 VPGASPKPLFAFRCGMVFLVD 203

RESULT 3
 US-09-107-532A-6355
 ; Sequence 6355, Application US/09107532A
 ; Patent No. 6583275
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn A Doucette-Stamm and David Bush
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 ; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
 ; NUMBER OF SEQUENCES: 7310
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
 ; STREET: 100 Beaver Street
 ; CITY: Waltham
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02354

COMPUTER READABLE FORM:
 ; MEDIUM TYPE: CD-ROM ISO9660
 ; COMPUTER: PC
 ; OPERATING SYSTEM: <Unknown>
 ; SOFTWARE: ASCII
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/107,532A
 ; FILING DATE: 30-Jun-1998
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 60/085,598
 ; FILING DATE: 14 May 1998
 ; APPLICATION NUMBER: 60/051571
 ; FILING DATE: July 2, 1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Ariniello, Pamela Deneke
 ; REGISTRATION NUMBER: 40,489
 ; REFERENCE/DOCKET NUMBER: GTC-012
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (781)893-8277
 ; TELEFAX: (781)893-8277
 ; INFORMATION FOR SEQ ID NO: 6355:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 344 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHEICAL: YES
 ; ORIGINAL SOURCE:
 ; ORGANISM: Enterococcus faecium
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (B) LOCATION 1...344
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 6355:
 ; US-09-107-532A-6355

Query Match 39.9%; Score 634; DB 4; Length 344;
 Best Local Similarity 40.5%; Pred. No. 2.6e-44;
 Matches 130; Conservative 75; Mismatches 106; Indels 10; Gaps 5;

QY 5 LKGVDM---KKKLLAGATLLSVATLAACSGSEGDALISMKGDTTEHOFTFEQVKSNS 61
 DB 2 ITGVQILMKKKSITLATSALAVITLAACS-GDTNKDIATMKGGTITVSPTFEAKLESS 60
 QY 62 AQQVLLNMTIOKVFEEQYSELDKDVDDTIAEEKOYGENORVLSOAGMTLETRKAQI 121
 DB 61 NQSLVQMIITIKYFNNKYGDKVTDKQYDAVDQAKNSLGTFFESQLEAAGYTRDTYKEYI 120
 QY 122 RTSKLVLAAYKVAEELTDEAYKKAFDEYTPDVTAQIIRLNNEDEKAEVLEKAKAEGAD 181
 DB 121 RNILAFKAGLK--AHVDITDDILKTAWKSPHVEAQIIRLSSEDEKADV-KRSADGDGD 177

QY 182 FAOLANDSTDEKTEKNGEITFPDSASTEVEQYKKAAPALDVGSDVITATGTOAYSS 241
178 FSKLAKKSDTDEKDGKGVKPDSTTTTIPAEVKEAFAFKLKDGEISDVITTTNPISYAT 237
QY 242 QYIVLTKTEKSSNIDOKELKVILLOKONDSFVOSITGELQANIKVQOAFQ 301
Db 238 EYVYKRVKNGNNGMDKRDQKDIATETKLSDNAFTTKVIGELKANVRIKIDAF 297
QY 302 NIFTYIGGDSSTSSSTSN 322
Db 298 NVLSAFT---TSSSTKDSSE 315

RESULT 4

US-09-071-035-182
; Sequence 182, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071.035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 182:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 342 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-071-035-182

Query Match 38.1%; Score 605.5; DB 4; Length 342;

Best Local Similarity 41.8%; Pred. No. 5.7e-42; Mismatches 133; Conservative 62; Indels 9; Gaps 5;

QY 10 MKKKLAGATLTLVATLTAACSKSGADLISMKGIVTTHOFEYQKSNPSAQOVLNM 69
Db 1 MKKKLITLAAGAMAVFLAACSSGK--DIATMKGSITITVDFFYNQIKEGSTSQAFSQM 58
QY 70 TIOKVFPEKQYSGEIDKEVDVDTIAEEKQ---YGENYQVLSQAGMTLETTRKAQIRTSKL 126
Db 59 VIKYVPEKYGKVDTKXIKNFDKAEQVEAGGKFSDLKQAGLTKETPKKOLKORAA 118
QY 127 VELAVKVAEELTDEAYKKAPEDEYTPDVAQIIRLNNEDKAKAEVLEKAAEGADFOLA 186
Db 119 YDAGLK--AHKITDEDLKTAAMASFHPEVEAQIIVASSEDPAKAV-KKEITDGGDFTKIA 175
QY 187 KDNSTDEKTEKNGEITFPDSASTEVEQYKKAAPALDVGSDVITATGTOAYSSQYIV 246

Db 176 KESTPTATKDGKIKFDSQATTPAEVKEAFAFKLKDGEVSEPIATNNQTYQTYVYV 235
QY 247 KLTKTEKSSNIDYKELKTVILLOKONDSFVOSITGELQANIKVQDAFONIFQ 306
Db 236 KMTKRAKGMKMPYEKIKKIAETKRLADQTVSKVISDELKAANVKIDAFKNALAG 295
QY 307 YIGGDSSTSSSTSN 322
Db 296 YM-QTESSSASSSEKE 310

RESULT 5

US-09-107-532A-6113
; Sequence 6113, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A. Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107.532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Arlinello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: CTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6113:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 303 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...303
; SEQUENCE DESCRIPTION: SEQ ID NO: 6113:
US-09-107-532A-6113

Query Match 37.1%; Score 588.5; DB 4; Length 303;

Best Local Similarity 40.5%; Pred. No. 1.2e-40; Mismatches 125; Conservative 69; Indels 7; Gaps 5;

QY 10 MKKKLAGATLTLVATLTAACSKSGADLISMKGIVTTHOFEYQKSNPSAQOVLNM 69
Db 1 MKKFLALAI-VLTGTLSCITNAGE-KTAVSYKGGTISFDVWDSIKKMGADSNVQOL 58
QY 70 TIOKVFPEKQYSGEIDKEVDVDTIAEEKQYGENYQVLSQAGMTLETTRKAQIRTSKVL 129

Db 59 IYQVPEKDYGDVSTKEIDSDYDTKQKLGDSFDSQLSAGYTEQTFKSDIKQSLAFOE 118
 QY 130 AKKVAEALITDEAVKKADETPDYTAQIIRLNNDKAKEVLEKAKAGADFAOLAKDNSTDEKRENGEITPDS 189
 Db 119 GKRC--HILTEDLKTAWSEFPEVEAOIIOVASDDAKDV--KKADGDDESKLAKDK 175
 QY 190 STDEKRENGEITPDSASTEVEEQYKKAFFALDVGVSDVITATGTQAYSQYIVTKLT 249
 Db 176 STDTTKEKGGKVKFPSTTTTVAPEKFAFKLKDQVSDVITSTASTTTTTEYYVKKV 235
 QY 250 KTEKSSNIDYKEKKTIVILTKONDSITFVOSITKELOANIKYKDAFOFIYQYIG 309
 Db 236 KKNONKNDKDYKKELEIATDTKLSDSTFONKVIQEVLDKANVKTIKDXFENVLSTFT- 294
 QY 310 GGDSSSSSS 318
 Db 295 -SDSSTASS 302

RESULT 6

US-09-071-035-184
 ; Sequence 184, Application US/09071035
 ; Patent No. 6448043
 ; GENERAL INFORMATION:
 ; APPLICANT: Gil H. Choi
 ; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
 ; NUMBER OF SEQUENCES: 496
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/071,035
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: A. Anders Brookes
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PB369P2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 309-8512
 ; INFORMATION FOR SEQ ID NO: 184:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 322 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-071-035-184

Query Match 35.4%; Score 562.5; DB 4; Length 322;
 Best Local Similarity 41.2%; Pred. No. 1,7e-38;

Matches 122; Conservative 58; Mismatches 107; Indels 9; Gaps 5;

QY 30 CSKSGEADLISKGDVITTEHOYEYKSNPSAQOVLNNTIOKVEFKQYSGELDDKEVD 89
 Db 1 CSSGSK--DIATKSGTTFVDFEYNOIKEOSTSQAFSQVYIKVEPEKGYDVTDXIXQ 58
 QY 90 DTIAEKKQ--GENTQVRLSQAGMTLETRKAOIRTSKLVLEAVKVAEALITDEAYKK 146
 Db 59 KNDEAKEOVEAOGKGFSDALKOAGLTKETFKOLKORAAVDAGLK--AHLKTIDEDLKT 116

QY 147 ADEVTDPDYTAQIIRLNNDKAKEVLEKAKAGADFAOLAKDNSTDEKRENGEITPDS 206
 Db 117 AMASFHEVEAOIIOVASDDAKAV--KKEITDGDFTKIAKESITPATKDGKTKFPS 175
 QY 207 ASTEVEEQYKKAFFALDVGVSDVITATGTQAYSQYIVTKLTKEKSSNIDYKEKIK 266
 Db 176 QATTVAEVKEAFAFKLKDGEVSEPIATNMQYOTTYVYVKKMKKAKGNDMKPYEKEIK 235
 QY 267 TVILTKONDSITFVOSITKELOANIKYKDAFOFIYQYIGGGSSSSSSSTNSNE 332
 Db 236 KIAETRTLDQYFVSVKVISDELKAAVNIKDAFKNALAGYM--QTESSASSSEKKE 290

RESULT 7

US-09-107-532A-4736
 ; Sequence 4736, Application US/09107532A
 ; Patent No. 6583275
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn A Doucette-Stamm and David Bush
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 ; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
 ; NUMBER OF SEQUENCES: 7310
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
 ; STREET: 100 Beaver Street
 ; CITY: Waltham
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02354
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: CD-ROM ISO9660
 ; COMPUTER: PC
 ; OPERATING SYSTEM: <Unknown>
 ; SOFTWARE: ASCII
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/107,532A
 ; FILING DATE: 30-Jun-1998
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 60/085,598
 ; FILING DATE: 14 May 1998
 ; APPLICATION NUMBER: 60/051571
 ; FILING DATE: July 2, 1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Ariniello, Pamela Denke
 ; REGISTRATION NUMBER: 40,489
 ; REFERENCE/DOCKET NUMBER: GTC-012
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (781)893-5007
 ; TELEFAX: (781)893-8277
 ; INFORMATION FOR SEQ ID NO: 4736:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 302 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHEICAL: YES
 ; ORIGINAL SOURCE:
 ; ORGANISM: Enterococcus faecium
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (b) LOCATION 1...302
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 4736:
 ; US-09-107-532A-4736

Query Match 28.4%; Score 450.5; DB 4; Length 302;
 Best Local Similarity 33.6%; Pred. No. 2.4e-29;

Matches 99; Conservative 61; Mismatches 130; Indels 5; Gaps 3;

QY 10 MKKKLAGATLLSVATLACSKSGEADLISKGDVITTEHOYEYKSNPSAQOVLNNT 69
 Db 12 LKQQLIMCLSAAILLFSGCSKSN--TIATFKGTLITVQDDYDNRKVSNNQTLKKM 69

```

QY      70  TIOKVEKGYSELDKEVDITIAEKKOYGENYQAVLSQAGMTLETRRAQIRTSKLVEL 129
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      70  IYFKAPDDVYGEKISSEKTEKXEYDQIKIGPNYKQOLKAVGQTEETFKKLFKQMLAFQY 129
QY      130  AVKVAEALITDEAKKAPDETPDYTAQIIRLNNDKAKVELEKKAQADPAQALAKON 189
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      130  GK--ANVLTQDKDLDTAWEKFEYVSTQIILFSTEEDAKKAKKAN--GGENSKLVQAY 186
QY      190  STDEKRENGETFPDSASTEVEPOYKKAFAVDVGSVDVITATGTOAYSSQYIYVTKL 249
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      187  GKNTLAKETDGMKNPSTNPEITPEYKKAFAFKAKNEVSDIIPVTPPTTYQOSYIVKAV 246
QY      250  KTEKSSNIDYKREKLTIVLQKONDSFVOSIIEKELQANIKVQDAFONIF 304
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      247  KKQDKSNKDKYKSELEKIALTEAKTLDTEFMDKTIRKVKKKDVTIKDYPVKNIF 301

```

RESULT 8

```

US-08-858-207A-508
; Sequence 508, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: No. 6348328el Compounds
; NUMBER OF SEQUENCES: 552
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858, 207A
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/017670
; FILING DATE: 14-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmil, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50475
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 508:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 74 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6348328e
US-08-858-207A-508

```

Query Match

```

Best Local Similarity 97.1%; Score 322; DB 4; Length 74;
Matches 66; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      10  MKKLLAGATLTSVTLAAGSGADLISMKGVITIEHQVEQYKSPSAQOVLNLM 69
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      1  MKKLLAGATLTSVTLAAGSGADLISMKGVITIEHQVEQYKSPSAQOVLNLM 60

```

```

QY      70  TIOKVEEK 77
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      61  TIOKVEEK 68

```

RESULT 9

```

US-09-134-001C-3811
; Sequence 3811, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: CTC-007
; CURRENT APPLICATION NUMBER: US/09/134, 001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3811
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3811

```

Query Match

```

Best Local Similarity 27.5%; Score 210; DB 4; Length 330;
Matches 95; Conservative 55; Mismatches 112; Indels 84; Gaps 20;

```

```

QY      23  SVATLAAC-SKGEAD--LISMK-GDV-----ITEHQVEQYKSNPSAQOVLNMTIOYV 74
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      19  SALLACGSGNATESNDNTLISKADYVADYKMKGEQI-ANTS-----FSIYLVNKV 72
QY      75  FEKQYSELDKKEVDITIAEKKQYQ--ENYQAVLSQAGMTLETRRAQIRTS---KLVEL 129
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      73  LADKYKQVDTKDIDKIDKEKQYQKQDFESMLKQOGMSLDYKEQKLSAYQOLLL 132
QY      130  AVKVAEALITDEAKKAPDETPDYTAQI--IRLNND-----KAKEVEKKAKE 178
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      133  DKVNVSDKEIKESKR-----TSHLLIKVKSQSDKEGLSDKKAKEKKEKIOKEV 182
QY      179  ---GADPAQALAKNSDDEKTEKENGGETFPDSASTEVEPOYKKAFAVDVGSVDV--TAT 234
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      183  EKPNKFGELAKKESDSSAKKQGSIGY-VINGQAVDSERKALFKLKEGEVSKVYKTDY 241
QY      235  GTQAYSSQYIYVLTQKTE---KSSNI-----DDYKREKLTIVLQKQ 274
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      242  G-----YHIIKADKETDENSEKSNIKQKLIBEKYQKKPKLLTDAYKELLKRYKDYKD 294
QY      275  ND--STFVOSIIEKELQANIKVQDAFONIFQYIGGDSSSSS 318
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      295  RDIKKAIEDSIIDP-----KITQOQOQ---OSOGSGSLTNSGS 330

```

RESULT 10

```

US-09-252-991A-17422
; Sequence 17422, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17422

```

```

; LENGTH: 447
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17422

Query Match
Best Local Similarity 20.1%; Score 147.5; DB 4; Length 447;
Matches 74; Conservative 72; Mismatches 140; Indels 83; Gaps 13;

QY 2 NTYLKGVDMKKLLAGATITLVSATLAACSGSGADLISM-----KGDVITEHOFEY 54
Db 17 NVKIKCNRLRPLALGALLCSFA-----HAENVPLDRVALVDVDMQSDLOQ 66
QY 55 QVK-----SNPSAQOVLNMTIQKVEK---QYSGELDKEDVD---TIA 93
Db 67 RLREHQTLLKRGAPLPPEHVLTOQVLERLIENIQOIGDRGIRISDELMQMGTTA 126
QY 94 EKKOYGENYQVYLSQAGMTLETIRKQIIRSKIV-ELAVKVAEALTDATYKAF----- 148
Db 127 QRNGMSLEQFQALTRGSLYADAREQVREMYISRVQRQVAREIQQVSEYKKNFLASD 186
QY 149 -----DEX-----TPDYTAQIIRLNNEDEKAEVLEKAEQADFAQLAKDNSTDE 193
Db 187 MGRIQLESEYRLANILITPVEAASDVYQAARQARELYQOLK-QGADFEQLAISAGD 245
QY 194 KTKENGELTFDASSTEVPEQVYKKAFFALDYGVSYITATGTQAVSSQYIYKLTKE 253
Db 246 NALE-GEIGWRKAA-QLPQPFDSMIGSLAVGVTPEVRPG-----GFILIKLEKRG 297
QY 254 KSSNIDYKKEKTLVILTKONDSFVQSIIGKELQANIKVDOAFQNFQYIGGSDS 313
Db 298 GSKMV-----RDEVVHRILKLPSELRSARETEKLAQKILYERIQSGEDP 341
QY 314 SSSSSSTISNE 322
Db 342 GELAKSEF 350

RESULT 11
US-09-328-352-7725
; Sequence 7725, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNIT FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7725
; LENGTH: 443
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7725

Query Match
Best Local Similarity 8.7%; Score 137.5; DB 4; Length 443;
Matches 78; Conservative 66; Mismatches 119; Indels 99; Gaps 16;

QY 7 GYDMKKLLAGATITLVSATLAACSGSGADLISMKGVDITEHOFEYQVKSNSAQVLT 66
Db 100 GIKPDEKSLNEAV--LKVAS-----QSGSKSLFAPOKDAIAPGT-YENLRS-----RIA 147
QY 67 LMTIQKVEKQYGS--ELDDKEVDVDTIAEKKQ-----YGENYQVYLSQA 110
Db 148 EELAINRLKQOQVMRIKISDDVDNFKLSPQGOALGNQAHVIMRISGDNRPQEVQYVA 207
QY 111 GMTLETRKQIIRTSKIVELAVKVAEALTDATYKAF-----DEYT 152
Db 208 ---KEVRSGLAQNNDL--NALKKLSIATYKVGADMGFRPLSDIPAEIARITPLQDDQT 262
QY 153 PD-----VTQAIIRLNNEDEK-----AKEVLE---KAKA 177

```

```

Db 263 TDLISVRDGVHVLKLERKQNEQKALVPQYQTHILIQSEVYSPENAKQIIDSITKRLK 322
QY 178 EGADFAQLAKDNSTDEKTKENGELTFDASSTEVPEQVYKKAFFALDYGVSYITATGTQ 237
Db 323 AGEDFATLATYASNDTGSARDGSLGWVTGMMVPERDK-----MGLPVEISE 373
QY 238 AYSQY--YVYKLTKEKSNIDVYKREKLTIVILQKONDSFVQSIIGKELQANIKV 295
Db 374 PFOIGFWHILQVTDKREKMT-HEYQERMAQIILERO-FNTEIDSMLEVRANAVVEI 431
QY 296 KD 297
Db 432 KD 433

RESULT 12
US-08-465-746-2
; Sequence 2, Application US/08465746
; Patent No. 5679768
; GENERAL INFORMATION:
; APPLICANT: Briles, David E
; APPLICANT: Yother, Janet L
; TITLE OF INVENTION: EPIITOPIC REGIONS OF PNEUMOCOCCAL SURFACE
; TITLE OF INVENTION: PROTEIN A
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Matzare, Ltd
; STREET: Suite 1203, 2001 Jefferson Davis Highway
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,746
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/048,896
; FILING DATE:
; APPLICATION NUMBER: US 07/656,773
; FILING DATE: 15-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/835,698
; FILING DATE: 12-FEB-1992
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 521-0378
; TELEX: LUKPAT WASHINGTON
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 619 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-465-746-2

Query Match
Best Local Similarity 8.5%; Score 135.5; DB 1; Length 619;
Matches 77; Conservative 54; Mismatches 116; Indels 69; Gaps 15;

QY 11 KKKLLAGATITLVSATLAACSGSGADLISMKGVDITEHOFEYQVKSNSAQVLTNMT 70
Db 3 KKKMI--LTSLASVAIL-----GAGFVASQPTVV-----RAESPVAS----- 38
QY 71 IQKVEKQYSGELD-----KEVDVDTIAEKKQYGENYQVYLSQAGMTLETIRK 118

```



```

Db      39  -QSKAEKDYDAAKDKANKKAVEDAQKALDPAKAAQK--YDEDDKTEEKAAL-----EK 92

Qy      119 AQIRTSKIVELAKKKAABELTDEAYKKAFEDYTPDYVAQIIRLNNEKKAEEVLEKAAE 178
      1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      93  A--ASSEEMDKAAVAAYQAYL---AYQAQTDKAADADAKMI----DEAKREEAKTK 141
      1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

Qy      179 -----GADFQALAKDNSTDTEKTEKENGGEIT--FDSASTVEPQYKKAFALDVGVGS 228
      1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

Db      142 FNTVRAMVYPEPQQLAEFKKKSEBAQKQKPELTLLKLEAKKKLEBAEKKALEAQKDAE 201
      1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

Qy      229 DVTATGTAQYASQYIYIVKLTTKTEKSSNIDY--KEKITYI--LQKQNDSTFVOSII 284
      1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

Db      202 EVAPQAKIAELEQVHRLQELKEIDSESEDEYAKEGFRAPLOSLDKAKKAKLSKLEELS 261
      1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

Qy      285 GK--ELQAAANIYKQD 298
      1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

Db      262 DKIDELDAETAKLEDD 277
      1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

```

01 RESULT 13
02 US-08-214-164-2
03 ; Sequence 2, Application US/08214164
04 ; Patent No. 5728367
05 ; GENERAL INFORMATION:
06 APPLICANT: BRILES, DAVID E.
07 APPLICANT: YOTHER, JANET L.
08 TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL PROTEIN
09 NUMBER OF SEQUENCES: 3
10 CORRESPONDENCE ADDRESS:
11 ADDRESSEE: Shoemaker and Mattare, Ltd
12 STREET: Suite 1203, 2001 Jefferson Davis Highway
13 CITY: Arlington
14 STATE: Virginia
15 COUNTRY: U.S.A.
16 ZIP: 22202-0286
17 COMPUTER READABLE FORM:
18 MEDIUM TYPE: Floppy disk
19 COMPUTER: IBM PC compatible
20 OPERATING SYSTEM: PC-DOS/MS-DOS
21 SOFTWARE: Patentin Release #1.0, Version #1.25
22 CURRENT APPLICATION DATA:
23 APPLICATION NUMBER: US/08/214,164
24 FILING DATE: 17-MAR-1994
25 CLASSIFICATION: 424
26 PRIOR APPLICATION DATA:
27 APPLICATION NUMBER: US 07/656,773
28 FILING DATE: 15-FEB-1991
29 CLASSIFICATION: 424
30 ATTORNEY/AGENT INFORMATION:
31 NAME: Berkstresser, Jerry W.
32 REGISTRATION NUMBER: 22,651
33 REFERENCE/DOCKET NUMBER: 6102-137
34 TELECOMMUNICATION INFORMATION:
35 TELEPHONE: (703) 415-0810
36 TELEFAX: (703) 521-0813
37 TELEX: LUKPAT WASHINGTON
38 INFORMATION FOR SEQ ID NO: 2:
39 SEQUENCE CHARACTERISTICS:
40 LENGTH: 619 amino acids
41 TYPE: amino acid
42 TOPOLOGY: linear
43 MOLECULE TYPE: protein
44 US-08-214-164-2
45
46 Query Match 8.5%; Score 135.5; DB 1; Length 619;
47 Best Local Similarity 24.4%; Pred. No. 0.0043;
48 Matches 77; Conservative 54; Mismatches 116; Indels 69; Gaps 15
49
50 11 KKLLAGATITLTSVATIACSGSGGADLSMKGDYITHQRYEYKSNPSAQOVLNMT 70
51 ::::||||| |:::|:|
52 Db 3 KKKM--DTSLASVAI-----GAGFVASQPVV-----RAESFPVS----- 38
53 71 IOKVEKGYSGELDD-----KEVDPIAEERKQYGENRQVSIQAGMTLETFR 118

```

Db	39	-GSAEKEDYDAKDAKKDAKNKRAVEDAOKALDDAKNAQOK--YDEDOQKTEERKAL---EK	92
Qy	119	AQIRTKIVELANKKVAEALTDDEAYKKAFFEDYTPDYAOIIRLNNDKAEVILEKAE	178
Db	93	A--ASBEMKAAAAVQAYL--AYQQAOTDKAKDAADKML---DEAKREEEKTK	141
Qy	179	-----GADFAQLAKDNSTDEKTEKENGEIT--FDSASTFEPBQYKKAFFALDDVGS	228
Db	142	FNTVRAMVYPEEPQJLAETKKESEEAOKORAPETLTKLEBAKALBEAEKKAIEAOKVDAE	201
Qy	229	DYITATGTOAVSSOYIVKLTTRKTEKSSNIDY--KEKLEKY---LTQKQNDFFYOSII	284
Db	202	EVAPQAKIAELNQVHRLPQELKEIDSESEDYAKEGFRAPLQSOILDAKKAKKLSLBEIS	261
Qy	285	GK--ELQANIKRYKDQ	298
Db	262	DKIDEDDAETAKLEDO	277

```

US-08-467-852A-3
: Sequence 3, Application US/08467852A
: Patent No. 5856170
:
: GENERAL INFORMATION:
: APPLICANT: BRILES, David E.
: APPLICANT: YOTNER, Janet L.
: APPLICANT: MCDANIEL, Larry S.
: TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL PROTEIN
: NUMBER OF SEQUENCES: 10
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: FROMMER LAWRENCE & HAUG LLP
: STREET: 745 Fifth Avenue
: CITY: New York
: STATE: NY
: COUNTRY: USA
: ZIP: 10151
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/467, 852A
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 435
:
: ATTORNEY/AGENT INFORMATION:
: NAME: KOMALSKI, Thomas J.
: REGISTRATION NUMBER: 32,147
: REFERENCE/DOCKET NUMBER: 454312-2064
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-588-0800
: TELEFAX: 212-588-0500
:
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 619 amino acids
: TYPE: amino acid
: STRANDEDNESS: n/a
: TOPOLOGY: linear
: MOLECULE TYPE: amino acid
: US-08-467-852A-3
:
Query Match 8.5%; Score 135.5; DB 2; Length 619;
Best Local Similarity 24.4%; Pred. No. 0.0043;
Matches 77; Conservative 54; Mismatches 116; Indels 69; Gaps 15
:
QY 11 KKLLIAGITLISVATTLAAGSKSGSGADLISMGKGVITIEHOFYEQVKSNSAQQVLLNMT 70
: |||: : ||| | : : : : :
Db 3 KKKMI-LTSLASVAIL-----GAGFVASQPTVV-----FAESPVAS-----38
:
QY 71 IQVFEKRYGSELD-----KEVDITAEKKRYGENTQRLVLSAQGMTLTRK 118
: |||: : ||| | : : : : :
Db 39 -QSKAEKRYDARKKAKNAKRAVEAOAKALDDAKAKAKK-VDEDCKTTEEKAAI-----EK 92

```

```

QY      11 KKKLAGAITLLSVATLAACSGKSEGADLISMKGIVTEHQFYEQVKSNPAAQVLLMNT 70
      |||:: :||| | | |:: :|: ::: :|: ::: :|:

```

```
Search completed: August 29, 2003, 10:37:34
Job time : 31 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2003, 10:36:19 ; Search time 61 Seconds
(without alignments)
722.029 Million cell updates/sec

Title: US-10-049-473a-2
Perfect score: 1588
Sequence: 1 SWTYLKGVMMKKLLAGAIT.....IFQYIGGGSSSSSSSTSNF 322

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 510680 seqs, 136781880 residues

Total number of hits satisfying chosen parameters: 510680

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA.*
1: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1541	97.0	313	11	US-09-769-787-163
2	916	57.7	213	9	US-09-765-272-34
3	212.5	13.4	320	8	US-08-781-986A-5192
4	138	8.7	670	11	US-09-298-523B-63
5	135.5	8.5	619	11	US-09-298-523B-63
6	131.5	8.3	711	9	US-09-298-523B-63
7	130.5	8.2	2478	11	US-09-815-242-5816
8	130.5	8.2	2478	9	US-09-815-242-5816
9	130	8.2	690	11	US-09-298-523B-61
10	129.5	8.2	701	11	US-09-298-523B-62
11	129.5	8.2	589	11	US-09-298-523B-14
12	129.5	8.2	691	11	US-09-298-523B-14
13	129.5	8.2	707	11	US-09-298-523B-14
14	129.5	8.2	929	11	US-09-298-523B-60
15	129.5	8.2	1576	15	US-10-037-182-16

16	129.5	8.2	1609	10	US-09-938-275-11	Sequence 11, Appl
17	129.5	8.2	1609	15	US-10-037-182-14	Sequence 14, Appl
18	129.5	8.2	1609	15	US-10-299-058-12	Sequence 12, Appl
19	123	7.7	2285	10	US-09-932-188A-2	Sequence 2, Appl
20	122.5	7.7	428	11	US-10-254-995-7	Sequence 7, Appl
21	122.5	7.7	439	11	US-09-056-019-37	Sequence 37, Appl
22	122.5	7.7	1766	10	US-09-742-096-3	Sequence 3, Appl
23	119.5	7.5	530	9	US-09-159-469-73	Sequence 73, Appl
24	119.5	7.5	530	9	US-09-159-469-73	Sequence 73, Appl
25	119.5	7.5	590	9	US-09-159-469-64	Sequence 64, Appl
26	119.5	7.5	590	9	US-09-159-469-64	Sequence 64, Appl
27	117	7.4	767	9	US-09-815-242-5899	Sequence 5899, Ap
28	117	7.4	767	9	US-09-815-242-13140	Sequence 13140, A
29	114.5	7.2	439	11	US-09-056-019-28	Sequence 28, Appl
30	114.5	7.2	2368	9	US-09-815-242-5635	Sequence 5635, Ap
31	114.5	7.2	2368	9	US-09-815-242-12389	Sequence 12389, A
32	114	7.2	431	12	US-10-254-995-3	Sequence 3, Appl
33	114	7.2	481	11	US-09-298-523B-6	Sequence 6, Appl
34	113.5	7.1	1167	9	US-09-815-242-11522	Sequence 11522, A
35	112.5	7.1	425	12	US-10-254-995-13	Sequence 13, Appl
36	112.5	7.1	437	11	US-09-056-019-32	Sequence 32, Appl
37	112	7.1	841	9	US-09-861-451A-30	Sequence 30, Appl
38	111.5	7.0	424	12	US-10-254-995-14	Sequence 14, Appl
39	111.5	7.0	437	11	US-09-056-019-29	Sequence 29, Appl
40	111.5	7.0	487	11	US-09-298-523B-9	Sequence 9, Appl
41	111	7.0	428	12	US-10-284-400-8	Sequence 8, Appl
42	111	7.0	429	12	US-10-284-400-16	Sequence 16, Appl
43	111	7.0	466	11	US-09-298-523B-4	Sequence 4, Appl
44	111	7.0	764	15	US-10-309-851-10	Sequence 10, Appl
45	111	7.0	1133	15	US-10-309-851-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-09-769-787-163
Sequence 163, Application US/09769787
Publication No. US20030091577A1
GENERAL INFORMATION:
APPLICANT: Microbial Techniques Limited
APPLICANT: Gilbert, Christophe FG
APPLICANT: Hansbro, Phillip M
TITLE OF INVENTION: Proteins
FILE REFERENCE: PWC/P21129WO
CURRENT APPLICATION NUMBER: US/09/769,787
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: GB 9816337.1
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: US 60/125164
PRIOR FILING DATE: 1999-03-19
NUMBER OF SEQ ID NOS: 388
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 163
LENGTH: 313
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-769-787-163

Query Match 97.0%; Score 1541; DB 11; Length 313;
Best Local Similarity 100.0%; Pred. No. 2.3e-107;
Matches 313; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 MKKLLGATITLVSATLAACSKSGEGADLSIMKGDVITHEQFQYKSNPSAQOVLNM 69
DB 1 MKKLLGATITLVSATLAACSKSGEGADLSIMKGDVITHEQFQYKSNPSAQOVLNM 60
QY 70 TIOQVEKQYSGSELDDEVDVDTIAEKKQYGENORVLSQAGMTLERRKQIRTSKVEL 129
DB 61 TIOQVEKQYSGSELDDEVDVDTIAEKKQYGENORVLSQAGMTLERRKQIRTSKVEL 120
QY 130 AVKVAEAEITDEAYKKAFDEYTPDVYQAIIRLNNEKAEVLEKAKAGADPAQLAKDN 189
DB 130 AVKVAEAEITDEAYKKAFDEYTPDVYQAIIRLNNEKAEVLEKAKAGADPAQLAKDN 189

Db 121 AVKVAEALTDYAEYKKAFFDEYTPDYTAQIIRLNNEKAKAEVLEKAKAGADFAQLAKDN 180
QY 190 STDEKTEKNGEITFDSASTEVPEQVKKAFALPDVGVSDVITATGTAQVSSOYIYKLT 249
Db 181 STDEKTEKNGEITFDSASTEVPEQVKKAFALPDVGVSDVITATGTAQVSSOYIYKLT 240
QY 250 KTKKSSNIDYKKEKLTIVITOKONDSTFVQSIITGELQANIKVKDAQFQNIETQYIG 309
Db 241 KTKKSSNIDYKKEKLTIVITOKONDSTFVQSIITGELQANIKVKDAQFQNIETQYIG 300
QY 310 GGDSSSSSTSNE 322
Db 301 GGDSSSSSTSNE 313

RESULT 2
US-09-765-272-34
Sequence 34, Application US/09765272
Patent No. US20020061545A1

GENERAL INFORMATION:

APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/961,083
FILING DATE: <unknown>
ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 34:

SEQUENCE CHARACTERISTICS:

LENGTH: 213 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 34:

US-09-765-272-34

Query Match 57.7%; Score 916; DB 9; Length 213;
Best Local Similarity 93.6%; Pred. No. 5,9e-61;
Matches 150; Conservative 0; Mismatches 5; Indels 8; Gaps 2;

QY 31 SKSEADADLSMGDVITEHOFYEOYKSNPSAOQVILNMTIOKVFKEQSGSELDDEKVD 90
Db 1 SKSEADADLSMGDVITEHOFYEOYKSNPSAOQVILNMTIOKVFKEQSGSELDDEKVD 60
QY 91 TIAEKKQYGENQVRLSOGMTLETTRKAQIRTSKVELAVKVAEALTDYAEYKKAFFDE 150
Db 61 TIAEKKQYGENQVRLSOGMTLETTRKAQIRTSKVELAVKVAEALTDYAEYKKAFFDE 120
QY 151 YTPDVTAQIIRLNNEKAKAEVLEKAKAGADFAQLAKDNSTDEKTEKNGEITFDSASTE 210
Db 151 YTPDVTAQIIRLNNEKAKAEVLEKAKAGADFAQLAKDNSTDEKTEKNGEITFDSASTE 210

Db 121 YTPDVTAQIIRLNNEKAKAEVLEKAKAGADFAQLAKDNSTDEKTEKNGEITFDSASTE 180
QY 211 VP-EQVKKAF-----LDVD 225
Db 181 VPGASPKRPLFAFRGMYFLDVD 203

RESULT 3
US-08-781-986A-5192

Sequence 5192, Application US/08781986A
Publication No. US2003005436A1

GENERAL INFORMATION:

APPLICANT: Charles Kunsch
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/781,986A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

ATTORNEY/AGENT INFORMATION:

NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248PP

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 5192:

SEQUENCE CHARACTERISTICS:

LENGTH: 320 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-781-986A-5192

Query Match 13.4%; Score 212.5; DB 8; Length 320;
Best Local Similarity 23.5%; Pred. No. 3.2e-08;
Matches 72; Conservative 70; Mismatches 107; Indels 57; Gaps 12;

QY 23 SVATLAAGSGSGAD-----LISMAGDVITEHOFYEOYKSNPSAOQVILNMTIOKVFKE 77
Db 14 SALLTAGC--GASATSKENTLILSKAGDVITADTKKIKGKDIOIANSFTEM-LNKILAD 70
QY 78 QYSELDKDEKVDPTIAEKKQY--ENYQVRLSOGMTLETTRKAQIRTSKVEL--LAVK- 132
Db 71 KYNNKNDKKIDIOIEKMKQYGGKDFEKALQOQGITADKYKENTRTAAYHKEILSDKI 130
QY 133 KVAEALTEA-----YKKAFFDEYTPDVTAQIIRLNNEKAKAEVLEKAKAGAD 181
Db 131 KIDSSEIKEDSKASHILIKVSKSDXEGLDDEKA-----KKAEEIQKVSNDPSK 183
QY 182 FAQLAKDNSTDEKTEKNGEITFDSASTEVPEQVKKAFALPDVGVSDVITATGTAQVSS 241
Db 184 FGEIANKESMDGSAKDELGY-VLKQGTDKFERALFKLKGVESEVYKS-----SF 236
QY 242 QYIVVILTKKTEKSSNIDYKKEKLTIVITOKONDSTFVQSIITGELQANIKVKDAQF 301
Db 237 GVHIIKADKPTDPSNSKOSLKEK-----VDQVQV-KNPKLLTDAYK 277

QY 302 NIFTQY 307
DB 278 DLKEY 283

RESULT 4
US-09-523B-63
; Sequence 63, Application US/09298523B
; Publication No. US20030059438A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298,523B
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 63
; LENGTH: 670
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523B-63

Query Match 8.7%; Score 138; DB 11; Length 670;
Best Local Similarity 20.6%; Pred. No. 0.032;
Matches 85; Conservative 56; Mismatches 138; Indels 134; Gaps 14;

QY 31 SKGSEADLISMKGDVITEHQFYEQVKNPSAOQVLLNMTQKVEKQY-----K 80
DB 181 SKKAEATRLBEIKTE---RKKAEEBAKRAESEEKKAEEKOKVDYTKGCKPKRRAKGV 237
QY 81 -----SELDKKEVDYTAIEE-----KKQYGENYQVLSQ----- 109
DB 238 SGELATPDKKKNKSSSVGEETLPSPSLMANANESOTHRKVDYDEYIKMLSEIQDLDR 297
QY 110 -----AGMTLETRKAQIRTSKLVELAY-----K 132
DB 298 RKHTQNNVNLINKLSAITKLYELSVLAKENSKKEELTSKKAELTAFAEQFKDTLKEPK 357
QY 133 KVAEAE-LTDEAYKKAFDEYTPD-----VTAQITRL 162
DB 358 KVAEAEKKEVEAKKADOKEDRNRNPTNYTKLELEASDVYKDAELELYEEANES 417
QY 163 NNEDKAEVLEKAKBADAQALAKONSTDEKTEKENGELTFDSASTEVPEQVYKAFAL 222
DB 418 RNEKIRQAKKEKSKAEATRIEIKITDOKKAE-----EEAKKAESEKKAEEK 470
QY 223 D-VDGVSADVTATGTAQYSSQYIVLKTKTEK--SSNIDY--KEKLTQYI---LTQKON 275
DB 471 QKDAEYALDA---KIAELEFYQRLKELEKELDESDSEYIKGIRAPLOSKLDTRKA 527
QY 276 DSTFVQSIIGK-----ELQANIKVKDAQFONIFTQYIGGDSSTSSSTSN 322
DB 528 KLSKLEELSPKIDELDAEIAKLEVQLKDAEGNNVVEAYFKEGLEKTTAEKKA 580

RESULT 5
US-09-882-774-1
; Sequence 1, Application US/09882774
; Publication No. US20030021795A1
; GENERAL INFORMATION:
; APPLICANT: Houston, Michael E.
; APPLICANT: Hodges, Robert
; TITLE OF INVENTION: use of Coiled-Coil Structural Scaffold to Generate
; FILE REFERENCE: 003592-007
; CURRENT APPLICATION NUMBER: US/09/882,774
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,892
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/213,387

; PRIOR FILING DATE: 2000-06-23
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 619
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-882-774-1

Query Match 8.5%; Score 135.5; DB 11; Length 619;
Best Local Similarity 24.4%; Pred. No. 0.044;
Matches 77; Conservative 54; Mismatches 116; Indels 69; Gaps 15;

QY 11 KKKLLGATILVATLAACSKGSEADLISMKGDVITEHQFYEQVKNPSAOQVLLNMT 70
DB 3 KKKMI---LTSLASVALI-----GAGFVASQPTV-----RAESPVAS----- 38
QY 71 IQKVEKQYSELD-----KEYVDYIAEKKQYGENYQVLSQAGMTLETRK 118
DB 39 -QSKAEKDYDAKKDAKNAKAVEDAQKALDDAKAAQK-YDEDKKTEKKAAL---EK 92
QY 119 AQIRTSKLVEIAKVAEALLTDEAYKKAFDEYTPVTAQITRLNNEKAEVLEKAKAE 178
DB 93 A---ASEEMKAVAAVQOAYL---AYQOATDKAKADAADKMI-----DEAKKREEKTK 141
QY 179 -----GADFAOLAKDNSTDEKTEKENGELTF--FDSASTEVPEQVYKAFALVDYGV 228
DB 142 FNTVIRAVYVEPEQALATKRSSEAKOKAPELTKLEAKAKIEAKKATEKOKYDAE 201
QY 229 DVITATGTAQYSSQYIVLKTKTEKSSNIDY--KEKLTQYI---LTQKONSTFVQSI 284
DB 202 EVAPQAKIAELENGVHRLFEQLEKIDESSEYAKEGFRAPLOSKLDKAKKLSKLELS 261
QY 285 GK---ELQANIKYKQD 298
DB 262 DKIDELDAEIAKLEDQ 277

RESULT 6
US-09-298-523B-3
; Sequence 3, Application US/09298523B
; Publication No. US20030059438A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298,523B
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 711
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523B-3

Query Match 8.3%; Score 131.5; DB 11; Length 711;
Best Local Similarity 20.7%; Pred. No. 0.11;
Matches 86; Conservative 55; Mismatches 137; Indels 137; Gaps 15;

QY 31 SKGSEADLISMKGDVITEHQFYEQVKNPSAOQVLLNMTQKVEKQY-----K 80
DB 221 SKKAEATRLBEIKTE---RKKAEEBAKRAESEEKKAEEKOKVDYTKGCKPKRRAKGV 277
QY 81 -----SELDKKEVDYTAIEE-----KKQYGENYQVLSQ----- 109
DB 278 SGELATPDKKKNKSSSVGEETLPSPSLMANANESOTHRKVDYDEYIKMLSEIQDL 337
QY 110 -----AGMTLETRKAQIRTSKLVELAY----- 131
DB 338 RKHTQNNVNLINKLSAITKLYELSVLAKENSKKEELTSKKAELTAFAEQFKDTLKEPK 397

OY 132 KVAEAE-LTDEAKKAFDEYTPD-----VTAQII 160
| | | | | : | | | : |
DB 398 KVAEAEKKAEAKKADOKEDRRNPTNTYKLELEIAESPVKKALELYKEBAN 457
| | | | | : | | | : |
OY 161 RLNDEKAEVLEKAKGADPAOLAKONSTDEKRENGGHTFDSASTVEPQYKAAAF 220
| | | | | : | | | : |
DB 458 ESRNEEKIKOKAEVESKKAATRIEK-IKTRKKA-----EAKRKESEKKAEE 509
| | | | | : | | | : |
OY 221 ALD-VDGVSDVITATGTQAVSSQYIYVLTREKTEK--SSNIDY-KEKKTAVI--LTQK 273
| | | | | : | | | : |
DB 510 AKQVDAEYVLEA---KIAELEVEYQRLKEKLEIDSESDYLTKEGLRAPLQSKIDTK 566
| | | | | : | | | : |
OY 274 QNDSTFVQSTIGK-----ELQANIVYKQDAFONITFYIGGSDSSSSSTSNE 322
| | | | | : | | | : |
DB 567 KAKLSKLEELSDKIDELDAETAKLEVOLKDAEGNNNEAYFEKLEKTTAEKKAEE 621
| | | | | : | | | : |

RESULT 7

US-09-815-242-5816
; Sequence 5816, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815, 242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ. ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5816
; LENGTH: 2478
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
US-09-815-242-5816

Query Match 8.2%; Score 130.5; DB 9; Length 2478;
Best Local Similarity 22.3%; Pred. No. 0.68;
Matches 63; Conservative 48; Mismatches 103; Indels 69; Gaps 12;
OY 28 AACSKSGADLISMKGDVTEH---QPEYQKSNPSAQOVLNMTIQKVFEPKQYGSSEL 83
| | | | | : | | | : |
DB 1569 AAEEGLNINNAATTTGDTTADTAQVQOLHANE-----VKPPAGK-- 1612
| | | | | : | | | : |
OY 84 DKEVDTTAEERKQYGEN---YORVLSQAGMTLETRKAQIRTSKIVELAVKVAEAEELT 140
| | | | | : | | | : |
DB 1613 --KELQAAADKKTQIQEPTNMQOELINDAKQEVDELENAKTN-----VDQSSSTREY 1664
| | | | | : | | | : |
OY 141 DEAYK-----KAFDEYTPDVTAQI-----IRLNEDK-----AKEVLEKAKAGAD 181
| | | | | : | | | : |
DB 1665 DNAYKGGAKKINAVKTFSEYKKDALAKIEDAYNAKYNVADNSNASTSEIAE-AKQKLAEE 1723
| | | | | : | | | : |

OY 182 FAOLAKDNSTDEKRENGGEITFDASTVEPQYKKAFAALVDGVSDVITATG--TQAYS 240
| | | | | : | | | : |
DB 1724 LKQPADQNVNQAATSKD-----DIEVOJHN-----DLDNINDYITPTKKSASAT 1766
| | | | | : | | | : |
OY 241 SQYIYVLTREKTEKSSNIDYKEKLEKTVIILTOKONSTFVQSTI 283
| | | | | : | | | : |
DB 1767 TDLAYADQKKNNISADJNATQDEKQOAIKQVDNQVQTALESII 1809
| | | | | : | | | : |

RESULT 8

US-09-815-242-12967
; Sequence 12967, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815, 242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ. ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12967
; LENGTH: 2478
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
US-09-815-242-12967

Query Match 8.2%; Score 130.5; DB 9; Length 2478;
Best Local Similarity 22.3%; Pred. No. 0.68;
Matches 63; Conservative 48; Mismatches 103; Indels 69; Gaps 12;
OY 28 AACSKSGADLISMKGDVTEH---QPEYQKSNPSAQOVLNMTIQKVFEPKQYGSSEL 83
| | | | | : | | | : |
DB 1569 AAEEGLNINNAATTTGDTTADTAQVQOLHANE-----VKPPAGK-- 1612
| | | | | : | | | : |
OY 84 DKEVDTTAEERKQYGEN---YORVLSQAGMTLETRKAQIRTSKIVELAVKVAEAEELT 140
| | | | | : | | | : |
DB 1613 --KELQAAADKKTQIQEPTNMQOELINDAKQEVDELENAKTN-----VDQSSSTREY 1664
| | | | | : | | | : |
OY 141 DEAYK-----KAFDEYTPDVTAQI-----IRLNEDK-----AKEVLEKAKAGAD 181
| | | | | : | | | : |
DB 1665 DNAYKGGAKKINAVKTFSEYKKDALAKIEDAYNAKYNVADNSNASTSEIAE-AKQKLAEE 1723
| | | | | : | | | : |
OY 182 FAOLAKDNSTDEKRENGGEITFDASTVEPQYKKAFAALVDGVSDVITATG--TQAYS 240
| | | | | : | | | : |
DB 1724 LKQPADQNVNQAATSKD-----DIEVOJHN-----DLDNINDYITPTKKSASAT 1766
| | | | | : | | | : |
OY 241 SQYIYVLTREKTEKSSNIDYKEKLEKTVIILTOKONSTFVQSTI 283
| | | | | : | | | : |
DB 1767 TDLAYADQKKNNISADJNATQDEKQOAIKQVDNQVQTALESII 1809
| | | | | : | | | : |

```

RESULT 10
US-09-298-523B-62
; Sequence 62, Application US/09298523B
; Publication No. US20030059438A1
GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUDOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; TITLE OF INVENTION: AND STRAINS THEREOF AND USES THEREFOR
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298,523B
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 62
LENGTH: 701
;
; TYPE: PRT
;
; ORGANISM: Streptococcus pneumoniae
US-09-298-523B-62

```

```

RESULT 11
US-09-298-523B-14
: Sequence 14, Application US/09298523B
: Publication No. US20030059438A1
: GENERAL INFORMATION:
: APPLICANT: BRILES et al.
: TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSCC), EPITOPIC REGIONS
: TITLE OF INVENTION: AND STRAINS THEREOF AND USES THEREFOR
: FILE REFERENCE: 454312-3140
: CURRENT APPLICATION NUMBER: US/09/298.523B
: CURRENT FILING DATE: 1999-04-23
: NUMBER OF SEQ ID NOS: 78
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 14
: LENGTH: 589
: TYPE: PRT
: ORGANISM: Streptococcus pneumoniae
US-09-298-523B-14

```

[illegible]

```
Db 454 AEESEKAAAEKQKVADEEYALEA--KIAELEVEQRLKEKEIDESDELYLKELR 510
QY 267 TVI---LTQKONDSTFVOSITGK-----ELQANIKVKDAQOFNIFTQYIGGSSSS 317
Db 511 APLQSKLDTKKAKLSKEELSDKIDELDAELAKLEFVQLKAGNNNEALFKEGLEKTTA 570
QY 318 STSNE 322
Db 571 EKKA 575

RESULT 12
US-09-298-523B-1
; Sequence 1, Application US/09298523B
; Publication No. US20030059438A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; TITLE OF INVENTION: AND STRAINS THEREOF AND USES THEREFOR
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298,523B
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 691
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523B-1

Query Match
Best Local Similarity 22.5%; Score 129.5; DB 11; Length 691;
Pred. No. 0.14; Indels 105; Gaps 16;
Matches 82; Conservative 53; Mismatches 125;

QY 42 MKGDVITEHQFYEQVKSNP5A--QOVL---LNMTIQKVEKQYGSLEDDKEVDDTIAEE 95
Db 269 VSGELATPDKKENDAKSSDSVGEETLPSPSLMANESQTEHR-----KDYDEYI--- 318
QY 96 KKOYGE-NYQRYLSQAGMTLETRKAOIRTSKIVELAV----- 131
Db 319 KKMISEIOLDRKHTONVNLINIKLSAIKTKYLYELSVLKENSKEELTSKKAELTAAFE 378
QY 132 -----KKVAEAE-LTDEAYKKAFDEYTPD----- 154
Db 379 QKPKDLPKPEKKVAEAEKKVEAKKADQKEEDRRNYPNTYKTLLEIAESDYKKA 438
QY 155 ----VTAQIIRLNEDKAEVLEKAKAEGADFAQLAKDNDSTDEKTKENGGEITFDSASTE 210
Db 439 EELVVEEANESNEKIKQAKKEVSKKAERLREK-IKTDKKA 490
QY 211 VPEQVYKKAAPALD-VGVSDVITATGTQAYSSQYIVKLTKTER--SSNIDY-KEKLK 266
Db 491 AEESEKAAAEKQKVADEEYALEA--KIAELEVEQRLKEKEIDESDELYLKELR 547
QY 267 TVI---LTQKONDSTFVOSITGK-----ELQANIKVKDAQOFNIFTQYIGGSSSS 317
Db 548 APLQSKLDTKKAKLSKEELSDKIDELDAELAKLEFVQLKAGNNNEALFKEGLEKTTA 607
QY 318 STSNE 322
Db 608 EKKA 612

RESULT 13
US-09-298-523B-2
; Sequence 2, Application US/09298523B
; Publication No. US20030059438A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; TITLE OF INVENTION: AND STRAINS THEREOF AND USES THEREFOR
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298,523B
```

```
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523B-2

Query Match
Best Local Similarity 20.9%; Score 129.5; DB 11; Length 707;
Pred. No. 0.15; Indels 113; Gaps 14;
Matches 77; Conservative 57; Mismatches 122;

QY 42 MKGDVITEHQFYEQVKSNP5A--QOVL---LNMTIQKVEKQYGSLEDDKEVDDTIAEEKQY 99
Db 269 VSGELATPDKKENDAKSSDSVGEETLPSPSLMANESQ-----TEHRKDY 314
QY 100 GENYQRYLSQ-----AGMTLETRKAOIRTSKIVELAV----- 131
Db 315 DEYIKKMLSEIQLDGRKHTPNVNLINIKLSAIKTKYLYELSVLKENSKEELTSKKAELT 374
QY 132 -----KKVAEAE-LTDEAYKKAFDEYTPD----- 154
Db 375 AAEQFKDLPKPEKKVAEAEKKVEAKKADQKEEDRRNYPNTYKTLLEIAESDYK 434
QY 155 ----VTAQIIRLNEDKAEVLEKAKAEGADFAQLAKDNDSTDEKTKENGGEITFDS 206
Db 435 VKAELELVKEEANESNEKIKQAKKEVSKKAERLREK-IKTDKKA 486
QY 207 ASTEVEPYKKAAPALD-VGVSDVITATGTQAYSSQYIVKLTKTER--SSNIDY-K 262
Db 487 AKKAESEKAAAEKQKVADEEYALEA--KIAELEVEQRLKEKEIDESDELYLK 543
QY 263 EKLKTYI---LTQKONDSTFVOSITGK-----ELQANIKVKDAQOFNIFTQYIGGDS 313
Db 544 EGRAPLQSKLDTKKAKLSKEELSDKIDELDAELAKLEFVQLKAGNNNEALFKEGLE 603
QY 314 SSSSTYSNE 322
Db 604 KTTAEKKA 612

RESULT 14
US-09-298-523B-60
; Sequence 60, Application US/09298523B
; Publication No. US20030059438A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; TITLE OF INVENTION: AND STRAINS THEREOF AND USES THEREFOR
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298,523B
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 60
; LENGTH: 929
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523B-60

Query Match
Best Local Similarity 8.2%; Score 129.5; DB 11; Length 929;
Pred. No. 0.21; Indels 125; Gaps 16;
Matches 82; Conservative 53; Mismatches 125;

QY 42 MKGDVITEHQFYEQVKSNP5A--QOVL---LNMTIQKVEKQYGSLEDDKEVDDTIAEE 95
Db 269 VSGELATPDKKENDAKSSDSVGEETLPSPSLMANESQTEHR-----KDYDEYI--- 318
QY 96 KKOYGE-NYQRYLSQAGMTLETRKAOIRTSKIVELAV----- 131
Db 319 KKMISEIOLDRKHTONVNLINIKLSAIKTKYLYELSVLKENSKEELTSKKAELTAAFE 378
```


Search completed: August 29, 2003, 10:45:39
Job time : 62 secs

[illegible]

```

RESULT 15
US-10-037-182-16
: Sequence 16, Application US/10037182
: Publication NO. US20030044899A1
: GENERAL INFORMATION:
: APPLICANT: Tytgvason, Karl
: APPLICANT: Dol, Masayuki
: TITLE OF INVENTION: Recombinant Laminin 10
: FILE REFERENCE: 99-274-F
: CURRENT APPLICATION NUMBER: US/10/037,182
: CURRENT FILING DATE: 2001-12-21
: PRIOR APPLICATION NUMBER: 60/257,449
: PRIOR FILING DATE: 2000-12-21
: PRIOR APPLICATION NUMBER: 60/279,282
: PRIOR FILING DATE: 2001-03-28
: NUMBER OF SEQ ID NOS: 36
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 16
: LENGTH: 1576
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-037-182-16

```

Query Match	8.2%	Score 129.5	DB 15	Length 1576
Best Local Similarity	20.9%	Pred. NO. 0.44		
Matches 65	Conservative 73	Mismatches 136	Indels 37	Gaps 13

```

OY      32 KGESEADISKSG---DVITHEOYEYGVKSNPSAOQVILNMTIOKVEKOGYSELDD-- 86
Db      1015 KLQELSESLIANIAGTGDEMYTDOAFEDRKL---EAEREVMDDLREKQDVKDDVQNLMBRLQ 10711
OY      87 EYDDTIAEEKOYGENOIVLSQAQMTLETBRAOI--RTSKVELAIVKVAEMLTDEAYK 145
Db      1072 RVNNTLSSQISRL--ONINMTIEEGNLALAEQARAHEVENTERLIEIASRLEKAAVAANVS 1130
OY      146 KAFDEYTPDYAQIIRLNN---EDRAKEVELEKAAEGADPAQOLAK---DNSTDE----- 193
Db      1131 VTQPESTDP-----NNMTULAEARKLAEKRHQEADLVYRAKKTINDSTENYLL 1183
OY      194 KTKENGEITTD--SASTEVPPOVKKAAPALVDG--VSDVITATGTOA---YSSOYIYVK 247
Db      1184 RTLAGEONTAEIEELNRYEODAKNIISODLEKQOARVAHEEKRAGDKRVEIYASVAOLISP 1243
OY      248 LTKRT--EKSSNIDYKEKLTVILITKONDNSTFVOSIIGKELOANIKYKVDQAFONIFT 305
Db      1244 LDSEITLNEANNIKMEALENLQGLIDQKAKDYEDLREDMRGRKELEVKNNLEKGTKEQGTAD 1303
OY      306 QYIGGDSSSS 316
Db      1304 QLLARADAKA 1314

```

BEST AVAILABLE COPY

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2003, 10:31:09 ; Search time 41 Seconds

(without alignments)
755.276 Million cell updates/sec

Title: US-10-049-473a-2

Perfect score: 1588
Sequence: 1 SMNYLKGVMMKKLAGAIT.....IFQYIGGSSSSSTSN 322

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

1: PIR_76:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1541	97.0	313	2	E95113
2	1538	96.9	313	2	D97982
3	512	32.2	308	2	E86840
4	400.5	25.2	291	2	AF1722
5	393.5	24.8	293	2	AC1352
6	375.5	23.6	299	2	A44858
7	373	23.5	299	2	S08083
8	367	23.1	299	2	A32313
9	357.5	22.5	292	2	S15269
10	328	20.7	294	2	AD1255
11	327.5	20.6	294	2	A11617
12	286	18.0	333	2	A83797
13	212.5	13.4	320	2	H89970
14	189.5	11.9	353	2	D97295
15	187	11.8	677	2	E75563
16	169.5	10.7	594	2	B64018
17	164.5	10.4	623	2	A64774
18	163.5	10.3	623	2	G90690
19	163.5	10.3	623	2	G85541
20	162	10.2	628	2	A10382
21	158.5	10.0	628	2	AG0558
22	158	9.9	619	2	E82141
23	156.5	9.9	297	2	S66102
24	155.5	9.8	282	2	E71662
25	155.5	9.8	336	2	D70113
26	151	9.5	282	2	H97809
27	151	9.5	331	2	AF3267
28	146	9.2	273	2	S52412
29	142.5	9.0	430	2	B83572

30	140.5	8.8	314	2	A87629	hypothetical prote
31	139.5	8.8	437	2	B70353	conserved hypothet
32	135.5	8.5	619	2	A41971	surface protein ps
33	135.5	8.5	619	2	A97887	surface protein ps
34	135	8.5	237	2	F70446	conserved hypothet
35	135	8.5	1526	2	A45605	mature-parasite-in
36	134	8.4	358	2	AB1746	hypothetical prote
37	133.5	8.4	888	2	T25713	hypothetical prote
38	132.5	8.3	313	2	F64069	survival protein s
39	131.5	8.3	623	2	G84985	peptidylprolyl iso
40	131	8.2	247	2	A96934	peptidyl-prolyl ci
41	131	8.2	624	2	PC6003	surface membrane p
42	131	8.2	979	2	B90601	hypothetical prote
43	130	8.2	329	2	G72319	hypothetical prote
44	130	8.2	2139	2	T18296	myosin heavy chain
45	129.5	8.2	1609	1	MMHDB2	laminin gamma-1 ch

RESULT 1

E95113 proteinase maturation protein, probable [imported] - Streptococcus pneumoniae (strain

C:Species: Streptococcus pneumoniae

C:Date: 03-Aug-2001 #sequence revision 03-Aug-2001 #text_change 03-Aug-2001

C:Accession: E95113

R:Retelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H

on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzappl

son, T.; Hickey, E.R.; Holt, I.E.

Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris

A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; M01D:21357209; PMID:11463916

A:Accession: E95113

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-313 <KUR>

A:Cross-references: GB:AE005672; PIDN:AK75102.1; PID:914972456; GSPDB:GN00164; TIGR:

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP0981

Query Match	Score	Length	DB 2:	Length	313:
Best Local Similarity	100.0%	Pred. No. 3.3e-74;			
Matches	313;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
QY	10	MKKLAGATTLTSLVATLAACSKSGEGADLSMKGDVITTEHOFEQYKSNPSAQOVLNM	69		
DB	1	MKKLAGATTLTSLVATLAACSKSGEGADLSMKGDVITTEHOFEQYKSNPSAQOVLNM	60		
QY	70	TIOKVFEEKYGESELDKEDVDITAEKKOYGENYQVLSQAGMTLEFRKQIRTSKVEL	129		
DB	61	TIOKVFEEKYGESELDKEDVDITAEKKOYGENYQVLSQAGMTLEFRKQIRTSKVEL	120		
QY	130	AVKVAABEITDEYKKAPEYTPDVTAQIIRLNNEKAKAEVLEKAKAEGADFOALKDN	189		
DB	121	AVKVAABEITDEYKKAPEYTPDVTAQIIRLNNEKAKAEVLEKAKAEGADFOALKDN	180		
QY	190	STDEKTEKNGEITFDSASTVEPQVKKAFALDQVSDVITATGQAVSSQYIVKLT	249		
DB	181	STDEKTEKNGEITFDSASTVEPQVKKAFALDQVSDVITATGQAVSSQYIVKLT	240		
QY	250	KTEKSSNIDDKKELKTVITLTKQNDSTFVQSTIGKELQANIKVQDAFONFTQYIG	309		
DB	241	KTEKSSNIDDKKELKTVITLTKQNDSTFVQSTIGKELQANIKVQDAFONFTQYIG	300		
QY	310	GGDSSSSSTSN 322			
DB	301	GGDSSSSSTSN 313			

RESULT 2

RESULT 5

ACI3352
post-translocation molecular chapterone homolog lmo2219 [Imported] - Listeria monocytogenes
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
A:Accession: ACI3352
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fehl, H.
D.: Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kretzl, J.; Kunz, M.; Kunst, F.; Kurapkai, G.; Madeno, E.; Maitournam, A.; Makarewicz, C.; Schuster, T.; Simoes, N.; Tierney, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.
A>Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: ACI3352
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-793 <GB>
A:Cross-references: GB:NC_003210; PIDN:CAD00297.1; PID:g16411689; GSPDB:GNO0177
C:Genetic source: strain EGD-e
A:Gene: lmo2219

Query Match
Best Local Similarity 32.4%, Score 393.5; DB 2; Length 293;
Matches 100; Conservative 77; Mismatches 15; Indels 17; Gaps 11;

OY 10 MKKLLAGATTLTSVATLAACSGEGADLSMKGVITTEHOFYEYQVKSNEQAQVILLM 69
| | | | : : : : | | : : : : : : : : : : : : : : : :
Db 1 MKKKLILGLVMMALFLSLAAC---GGGDVDVKTDSGVTKDELDAMDKDGSEFV-QQL 56
| | | | : : : : | | : : : : : : : : : : : : : : : :
OY 70 TIOKVFEKYQVSELEDDKEVDVTIAEKKQVEENQVRLSQAGMLTRKAQIRSKIVEL 129
| | | | : : : : | | : : : : : : : : : : : : : : : :
Db 57 TFEKILGDGY--KVSDEVDDKRFNEYKSYQGDPSAVLTOSGLEKFFSOLKYNLLVQ 114
| | | | : : : : | | : : : : : : : : : : : : : : : :
OY 130 AVKVAEAELTDENVKRAAFDEYTPDYTAQIIRLNNEBKAKVLEKAEAGADPQALAKDN 189
| | : : : : | | : : : : | | : : : : : : : : : : : :
Db 115 ATE--ANPDISDTKLKTYETWQDPITYSHLVADENKAKVEBQKLK-DGEKFPADLKKEY 171
| | | | : : : : | | : : : : : : : : : : : : : : : :
OY 190 STDEKTENGEEITFDASSTVEPEOVKKAFAALDVGDVITATGTQAYSSQYYIVKLT 249
| | | | : : : : | | : : : : : : : : : : : : : : : :
Db 172 STDPAATKONGOLA-PFGPKMDPAFEKAAVALKNKG--DISAPVKTQ-YGIIMD 225
| | | | : : : : | | : : : : : : : : : : : : : : : :
OY 250 KKTERSSNIDYKREKLTIVILTOKONSTFVQSIIIGELQAANKVDAQNFOTYIG 309
| | : : : : | | : : : : | | : : : : : : : : : : : :
Db 226 KPATRTTFEKQ-KRAVNASYL-ESQLTTENNQTKLRKEDKANVVEDKDLDAFKDFDG 283
| | : : : : | | : : : : | | : : : : : : : : : : : :
OY 310 GGDSSSSSS 318
| | | | : : : : | | : : : : : : : : : : : : : : : :
Db 284 SSSSDSDSS 292
| | | | : : : : | | : : : : ~~~~~~

RESULT 6

A44858
probable protein export protein prtm precursor - Lactobacillus paracasei (strain NCDO 15)
C:Species: Lactobacillus paracasei
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
A:Accession: A44858
R:Holt, A.; Naes, H.
J. Gen. Microbiol. 138, 1353-1364, 1992
A>Title: Cloning, sequencing and expression of the gene encoding the cell-envelope-associated reference number: A44858; MUID:92381481; PMID:1512565
A:Accession: A44858
A:Molecule type: DNA
A:Residues: 1-299 <HOI>
A:Cross-references: GB:M3946; NID:g149580; PIDN:AAA5247.1; PID:g149581
A:Experimental source: strain NCDO 151
A>Note: sequence extracted from NCBI backbone (NCBIN:112261, NCBITP:112262)
C:Keywords: protein export
F:1-26/Domain: signal sequence #status predicted <SIG>

Query Match	23.6%	Score 375.5	DB 2	Length 299	
Best Local Similarity	31.2%	Pred. No. 4.9e-13			
Matches	100	Conservative	62	Mismatches 129	Indels 29
					Gaps 9
QY	10	MKKK-----LAGAATLLSVATLAAACSGSGADLISMKGDVITEHOFEOVSNPSAQ	64		
DB	1	MKKKMRLLKVLASTATFALLL--LSGCGSNQADQVATVYSGGKVTESFYEKLKOSPPTTK	58		
QY	65	VLLNMTIQKPEKQYGESELDDEKVEDDTIAEKKQYGENYORVLSQAGMTLETAKQIRTS	124		
DB	59	MLAMMLLYRALNHAYGVSSTKTYNADYDSKYQYGENPFAFLSQNFSKSSKESIRTN	118		
QY	125	KLVELA---YKKVAEAEITDEAYKKAARDEYTPDTAQIIRLNEDDKAEVLEKAKAGAD	181		
DB	119	FLSEVALKLLKKVSESOL-----KAVKKTQPKYTVYHILTSDEDTAKQVYISDL--AAGKD	172		
QY	182	FAQLAKDNDSTDEKTEKNGEITPDASTEVEPYQKKAFAALDVGVSVDVTATGTQAYSS	241		
DB	173	FATLAKTDSIDTATKNDGKISFELNNKTLIDATPKDAAYKLLK---NGDYTGTPPVKV--TD	227		
QY	242	QYIYVKLTKTEKSSNIDDYKEKIKYIILQKQNDSTFVQSIIGKELOAANIKYKQDAFQ	301		
DB	228	GVEYIKMINHPAKGT--FTSSKKALTASVYAKWSDSSIMORVISOVLKNOHVITIKDKDLA	286		
QY	302	NIFTYIGGGSSSSSSSTSN 321			
DB	287	DAL-----DSYKKPATTN 299			
RESULT 7					
S08083					
probable protein export protein prtm precursor - Lactococcus lactis subsp. cremoris (
C:Species: Lactococcus lactis subsp. cremoris					
C:Date: 30-Jun-1992 #sequence, revision 30-Jun-1992 #text, change 20-Sep-1999					
C:Accession: S08083					
R:Kiwaiki, M.; Ikehara, H.; Shimizu Kadote, M.; Hirashima, A.					
submitted to the EMBL data library, January 1989					
A:Description: Molecular characterisation of a cell wall associated proteinase gene o					
A:Reference number: S08082					
A:Accession: S08083					
A:Molecule type: DNA					
A:Residues: 1-299 <KIM>					
A:Cross-references: EMBL:X14130; NID:947197; PIDN:OAA32349.1; PID:947198					
C:Genetics:					
A:Genome: plasmid					
C:Keywords: protein export					
F:1-26/Domain: signal sequence #status predicted <SIG>					
Query Match	23.5%	Score 373	DB 2	Length 299	
Best Local Similarity	31.7%	Pred. No. 6.7e-13			
Matches	97	Conservative	59	Mismatches 128	Indels 22
					Gaps 8
QY	10	MKKK-----LAGAATLLSVATLAAACSGSGADLISMKGDVITEHOFEOVSNPSAQ	64		
DB	1	MKKKMRLLKVLASTATFALLL--LSGCGSNQADQVATVYSGGKVTESFYEKLKOSPPTTK	58		
QY	65	VLLNMTIQKPEKQYGESELDDEKVEDDTIAEKKQYGENYORVLSQAGMTLETAKQIRTS	124		
DB	59	MLAMMLLYRALNHAYGVSSTKTYNADYDSKYQYGENPFAFLSQNFSKSSKESIRTN	118		
QY	125	KLVELA---YKKVAEAEITDEAYKKAARDEYTPDTAQIIRLNEDDKAEVLEKAKAGAD	181		
DB	119	FLSEVALKLLKKVSESOL-----KAAKTYTPKTYTVYHILTSDEDTAKQVYISDL--AAGKD	172		
QY	182	FAQLAKDNDSTDEKTEKNGEITPDASTEVEPYQKKAFAALDVGVSVDVTATGTQAYSS	241		
DB	173	FATLAKTDSIDTATKNDGKISFELNNKTLIDATPKDAAYKLLK---NGDYTGTPPVKV--TD	227		
QY	242	QYIYVKLTKTEKSSNIDDYKEKIKYIILQKQNDSTFVQSIIGKELOAANIKYKQDAFQ	301		
DB	228	GVEYIKMINHPAKGT--FTSSKKALTASVYAKWSDSSIMORVISOVLKNOHVITIKDKDLA	286		

A: Authors; Kreft, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Maenuo, E.; Maltournam, A.; Matouren, A.; Vazquez-Boland, J.A.; Ok, C.; Schlutener, T.; Simoes, N.; Tierrez, A.; Voss, H.; Wehländ.

DB 36 AILDELVQRRTLSQAAVSLELTFQEELTDELETLFNSVLSELDDEMLALSNFGGLIENK 118
OY 118 KAQIRTSKLVELAARKVA--EAELTDEAYKKAFDEYTP--DYTAQIIIRLNEDKAKEVL 172

Db 116 EEFVNEYLPLVLEKLAEGVEITDEEKQAFDENRSLIEVASHILVEDEETAEVL 175
QY 173 EKAKAGADFAOLAKONSTDEKTEKNGEITFDSASTVEQVKKAFALDVGVSDVT 232
Db 176 DRLER-GDFAELAEISVSPSAEANNNGDIFGFGKDWPE-FEERAFMEIDESEPE 233
QY 233 ATGQVASSQYIVKLTKEKSSNIDYKEKLTAVILTKONDSFVOSIIGKELQAN 292
Db 234 ST---YG--YHIIIV---TDRKDSYELEEKIHDTLMNER---SRTQEVRLDLAQND 281
QY 293 IKVQQAQONIF-----TOYIGGDSSTSSSTSTNE 322
Db 282 INVLDQFEGFLDLPDAPPEVDPEIDGEDASDAEDQAE 321

RESULT 13

H89970
hypothetical protein prsa [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
R:Accession: H89970
R:Kucoda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-U, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratsugu, K.;
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: H89970
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-320 <KUR>
A:Cross-references: GB:BA000018; PID:g13701634; PIDN:BA042927.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: prsa

Query Match 13.4%; Score 212.5; DB 2; Length 320;
Best Local Similarity 23.5%; Pred. No. 0.00019;
Matches 72; Conservative 70; Mismatches 107; Indels 57; Gaps 12;

QY 23 SVATLAACSGSEGD-----LISMKGDTTEHQFEQVKNPSAQOVLNMTIOKVFEP 77
Db 14 SALLIGAC--GASADSCKENTLISKAGVTVADIMKKIGKQIANTSTEM-LNKILAD 70
QY 78 QYSELDDKEVDVDTIAEKKQY--ENTQVLSQAGMTLETRKAQIRSKLVE--LAVK- 132
Db 71 KYKNVNDKIKIDEQLEKMQYQYGRKDFEKAQOOGCLFADKYKEMLRTAAYHKELSDKI 130
QY 133 KYAEHELDEA-----YKAFDEYTPDVTRQIIRLNNEKAKAVLEKAKAEGAD 181
Db 131 KISDEIKESDKKASHIILIKYKSKSDKGLDDKEA-----KQKAEIOREVSNDPEK 183
QY 182 FAQLAKNDSTDEKTEKNGEITFDSASTVEQVKKAFALDVGVSDVTATGQVASS 241
Db 184 FGEIKAKESMDGSAKKQGELOY-VLKGOTDKDFEKAFLKLDGEVSEVVK-----SF 236
QY 242 QYIVKLTKEKSSNIDYKEKLTAVILTKONDSFVOSIIGKELQANIKVQDAQ 301
Db 237 GYHIIKAKPPTDFNSEKOSLKEKL-----VDQKVQ-KNPKLLTDAYK 277
QY 302 NIFQY 307
Db 278 DLAKY 283

RESULT 14

D97295
parvulin-like peptidyl-prolyl isomerase, PRSA B. subtilis ortholog [imported] - Clostridium
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: D97295
R:Rolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: D97295
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-353 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK81151.1; PID:g15026287; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC3215

Query Match

Best Local Similarity 11.9%; Score 189.5; DB 2; Length 353;
Matches 93; Conservative 59; Mismatches 113; Indels 87; Gaps 21;

QY 6 KGYDMMK-----KLLAGAITLSVATLAACSGSGADLSMK-----GDVITTEHF 52
Db 17 RGIKMSAKQIATALLVGMFTFSAV---GCSVKEKRPAINSKVAVIYGNQITRGEI 72
QY 53 -----YQVKS-----NPSA-----QOVLNMTIOKPEKQYSGELDDK--- 86
Db 73 DKLAKGVQLKSOYQDSYERKNEAVAAALKQKEQILTSIDQKITLKR---AKDQKITL 129
QY 87 -----EVDVITAEKKQY--GENYQVLSQAGMTLETRKAQIR---TSKIVELAVK 133
Db 130 TRDEIKTVNDVYVAYQYQEFTESEKFSOLSQYGYVAFEFKQOLKNRAISNKLIDQVVD 189
QY 134 VAEELTDEAYKKAPEYTPDV-----QIIRLNNEKAKAVLEKAKAEGADFAQ 184
Db 190 V--KVSDEAKKYYDSHKNSTQSPNTVHLAIL-VKTEKEKAV--KARIDKGEFAT 243
QY 185 LAKNDSTDEKTEKNGEITFDSASTVEQVKKAFALDVGVSDVTATGQVASSQ 242
Db 244 VAKQVSTD-GSKEKGDDLDIQENDSNYDKTFMAALKLINDQVSAFVTFQG----- 295
QY 243 YIIVKLTKEKTE-KSSNIDYKEKLTAVILTKONDSFVOSIIGKELQANI 293
Db 296 WHVIRKIKKTEYKPVKDFNSVKDIDKQTVLSTRO--KSVQKTLKWKESQANI 345

RESULT 15

E75563
probable peptidyl-prolyl cis-trans isomerase, Ppic family - Deinococcus radiodurans (s

C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: E75563
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: E75563
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-677 <WHI>
A:Cross-references: GB:AE001870; GB:AE000513; NID:g6457732; PIDN:AAF09662.1; PID:g645
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0071
A:Map position: 1

Query Match

Best Local Similarity 11.8%; Score 187; DB 2; Length 677;
Matches 91; Conservative 52; Mismatches 125; Indels 82; Gaps 16;

QY 11 KKKLLAGAITLSVATLAACSGSGADLSMKGDVITTEHQFEQVKN 59
Db 27 KRTAVNGLLIGLSVLYVAGMAYQFPNPGVSLFRNROGTFYIRVNGAVTAEOLESARRAS 86
QY 60 P-----SAQOVLNMTIOKVFEPKQYSGELDDKVD--PPIAEKKQYGEN----- 102


```

b      87  PLNDPEPSVLADDAKVIYISQ
y      103 -YQVLSOAGTLETREKAOIRF
y      147 AMTDAISOENGLTDSFRTQVRI
b      152 TPDVYAOIIRLNNSDRAKEVLEI
y      205 EPQYGRIVYTDKAKOSLLA
b      207 --ASTVEPYQKKAFAALDYDG
y      262 PVLAQALPTAVGEAFAFLKNGG
b      259 DDYKEKIKTIVLTQKKDSFVFF
y      316 D-----VVSAYRQOKN-----

```

```

search completed: August 29, 2003, 1
job time : 44 secs

```

BEST AVAILABLE COPY

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2003, 10:21:29 ; Search time 23 Seconds

(without alignments)
658.374 Million cell updates/sec

Title: US-10-049-473a-2

Perfect score: 1588

Sequence: 1 SMTYLGKGYDMKKLAGAIT.....IFTGYIGGSSSSSSSTNE 322

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	375.5	23.6	299 1	PRTM_LACPA
2	367	23.1	299 1	PRTM_LACLC
3	357.5	22.5	292 1	PRSA_BACSU
4	171	10.8	622 1	PPID_HAEIN
5	164.5	10.4	623 1	PPID_ECOLI
6	156.5	9.9	297 1	YACD_BACSU
7	155.5	9.8	282 1	PLP_RICPR
8	151	9.5	283 1	PLP_RICCN
9	146	9.2	273 1	CBP2_CAMJE
10	146	9.2	621 1	PPID_BUCAP
11	132.5	8.3	313 1	SUPA_HAEIN
12	131.5	8.3	623 1	PPID_BUCAT
13	129.5	8.2	1609 1	LMG1_HUMAN
14	127.5	8.0	673 1	PKNX_STRCO
15	127.5	8.0	1790 1	USOI_YEAST
16	126.5	8.0	2116 1	MYSD_DICDI
17	123.5	7.8	544 1	THS_MERTL
18	122.5	7.7	310 1	MTSA_STRGC
19	121.5	7.7	845 1	NFM_RAT
20	120.5	7.6	555 1	LEB8_DAVXA
21	120	7.6	708 1	RPSD_MYXCA
22	120	7.6	1433 1	RESE_CHICK
23	120	7.6	1959 1	MYHB_CHICK
24	119	7.5	1978 1	MYHB_CHICK
25	118.5	7.5	848 1	NFM_MOUSE
26	118.5	7.5	1841 1	CC12_SCHPO
27	118	7.4	1290 1	LA50_SCHPO
28	118	7.4	1535 1	LM1L_CAEEL
29	118	7.4	1961 1	MYH9_RAT
30	117	7.4	843 1	YMSI_YEAST
31	116.5	7.3	1528 1	SPG2_STRGO
32	116.5	7.3	1528 1	SPG2_STRGO
33	116.5	7.3	1875 1	MLP1_YEAST

34	116	7.3	1014 1	NEBL_HUMAN	076041 homo sapien
35	116	7.3	1300 1	Y338_MYCPN	P75296 mycoplasma
36	116	7.3	1960 1	MYH9_HUMAN	P35579 homo sapien
37	116	7.3	2469 1	TEG0_HSVSA	001056 herpesvirus
38	115.5	7.3	274 1	OS4_BORBU	004851 borrelia bu
39	115.5	7.3	679 1	YKR9_YEAST	P34237 saccharomyc
40	115	7.2	498 1	HMMR_RAT	P97779 rattus norv
41	115	7.2	1976 1	MYH9_HUMAN	P35580 homo sapien
42	115	7.2	2869 1	RBP1_PLAVB	000798 plasmodium
43	114.5	7.2	778 1	TOP1_CANAL	000313 candida alb
44	114	7.2	273 1	OS45_BORBU	009089 borrelia bu
45	114	7.2	1637 1	MRSP_STRAU	P80544 staphylococ

ALIGNMENTS

```

RESULT 1
ID      PRTM_LACPA      STANDARD:      PRT:      299 AA.
AC      002473;
DT      01-JUL-1993 (Rel. 26, Created)
DT      01-JUL-1993 (Rel. 26, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Protease maturation protein precursor.
GN      PRTM.
OS      Lactobacillus paracasei.
OC      Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC      Lactobacillus.
OX      NCBI_TaxID=1597;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      SRRAIN-NCDO 151;
RX      MEDLINE=92361461; PubMed=1512565;
RA      Holck A., Naes H.;
RT      "Cloning, sequencing and expression of the gene encoding the cell-
RT      envelope-associated proteinase from Lactobacillus paracasei subsp.
RT      paracasei NCDO 151."
RL      J Gen Microbiol 138:1353-1364(1992).
CC      -1- FUNCTION: THIS PROTEIN IS ESSENTIAL FOR PRODUCTION OF ACTIVE FORMS
CC      OF THE SERINE PROTEINASE LOCATED IN OR SECRETED FROM THE CELL
CC      ENVELOPE.
CC      -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC      (Probable).
CC      -1- SIMILARITY: BELONGS TO THE PRIC/PARVULIN FAMILY OF ROTMASES.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      EMBL: M83946; AAA25247.1; -
CC      PIR: A44858; A44858.
DR      InterPro: IPR000297; Rotomase.
DR      Pfam: PF00659; Rotomase; 1.
DR      PROSITE: PS01096; PPIC_PP1ASE_1; 1.
DR      PROSITE: PS01098; PPIC_PP1ASE_2; 1.
DR      PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW      isomerase; Rotomase; Membrane; Lipoprotein; signal.
FT      SIGNAL      1..23      POTENTIAL.
FT      CHAIN      24..299      PROTEASE MATURATION PROTEIN.
FT      LIPID      24..24      N-ACYL DIGLYCERIDE (BY SIMILARITY).
FT      DOMAIN      144..236      PPIC.
SQ      SEQUENCE      299 AA; 33145 MW; B9CF1EAB2F8C5867 CRC64;
OY      10 MKKK-----LLAGAILTLVATLLAAGSKSGADLIAMKGDVTEHQFYEQVSNPFAAQ 64
  
```

```

Db      1 MKKKMKRLKVLASTATALL--LSGCOSNQADQKATYSGGKVTESNFKELKQSPPTKT 58
QY      65 VLNMNTIOKVEFEKQYSELDKEDVDITAEKKQYGENYORVLSQAGMTLETRKAQIRTS 124
Db      59 MLANMLIYRALNHAYGSAVSTKTVNDAYDSYKQOYGENDFALPSQNGFRRSSEKELRTN 118
QY      125 KLVELA---VKVAEELDEAYKKAFFDEYTPDTAQIIRLNNEDEKAEVLEKAKAGAD 181
Db      119 FLSEVALKRLKLVSEQL-----KAAMKTYQPKVTYQHILTSDEDTAKQVLSDL-AAAGKD 172
QY      182 FAOLAKNSTDEKTEKNGEITPDSASTVEPQVKKAFALVDGVSIVTATGTQAYSS 241
Db      173 FATLAKTDSIDTATKNGKISFESNNKTLDATPKDAAYKLK-----NGDYTOTPVKV-TN 227
QY      242 QYIYVLTKEKTESNIDYEKELKTVILTKONDSFTVOSIIGKELQANIKRYKQDAFO 301
Db      228 GVEYIKMINHPAKGT-FTSSKRALTAASYAKWSRDSIMQVISOVLKNOHVITIKDKDLA 286
QY      302 NIFQYIGGDDSSSSSSSTSN 321
Db      287 DAL-----DSYKKPATYN 299

```

RESULT 2

```

PRTM_LACLC STANDARD; PRT; 299 AA.
AC      P14308; P15294;
DT      01-JAN-1990 (Rel. 13, Created)
DT      01-JAN-1990 (Rel. 13, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Protease maturation protein precursor.
PRTM.
OS      Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OC      Plasmid pSK11, Plasmid pLP763, and Plasmid pMW05.
OX      Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
RN      NCBI_TaxID=1359;
RP      SEQUENCE FROM N.A.
RC      STRAIN-SK11; PLASMID-PK11L;
RX      MEDLINE=89213972; PubMed=2496115;
RA      Vos P., van Asselodonk M., van Jeveren F., Slezén R., Simons G.,
de Vos W.M.;
RT      "A maturation protein is essential for production of active forms of
Lactococcus lactis SK11 serine proteinase located in or secreted from
the cell envelope."
RL      J. Bacteriol. 171:2795-2802(1989).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN-WG2;
RX      MEDLINE=89213971; PubMed=2708318;
RA      Haandrikman A.J., Kok J., Laan H., Soemilto S., Ledebor A.M.,
Koulogis W.N., Venema G.;
RT      Identification of a gene required for maturation of an extracellular
Lactococcal serine proteinase."
RL      J. Bacteriol. 171:2789-2794(1989).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN-NCD0 763; PLASMID-PLP763;
RA      Kiwaki M., Ikemura H., Shimizu Kadota M., Hirashima A.;
RL      Submitted (JAN-1989) to the EMBL/GenBank/DBD databases.
RN      [4]
RP      SEQUENCE OF 294-299 FROM N.A.
RC      STRAIN-SK11, and WG2; PLASMID-PK11L, and pMW05;
RX      MEDLINE=90343335; PubMed=2166472;
RA      Haandrikman A.J., van Leeuwen C., Kok J., Vos P., de Vos W.M.,
Venema G.;
RT      Insertion elements on Lactococcal proteinase plasmids."
RL      Appl. Environ. Microbiol. 56:1890-1896(1990).
CC      -1- FUNCTION: THIS PROTEIN IS ESSENTIAL FOR PRODUCTION OF ACTIVE FORMS
OF THE SERINE PROTEINASE LOCATED IN OR SECRETED FROM THE CELL
ENVELOPE. PROBABLE PRASE.
CC      -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor

```

```

CC      (Probable).
CC      -1- SIMILARITY: BELONGS TO THE PIC/PARVULIN FAMILY OF ROTAMASES.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb.ch/announce/
or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; J04962; AAA03532.1; -
DR      EMBL; M26694; AAA60395.1; -
DR      EMBL; X14130; CA132349.1; -
DR      EMBL; M37395; AAA25209.1; -
DR      PIR; A32314; A32313.
DR      PIR; S08083; S08083.
DR      InterPro; IPR000297; Rotamase.
DR      Pfam; PF00639; Rotamase; 1.
DR      PROSITE; PS01096; PIC-PRASE_1; 1.
DR      PROSITE; PS50198; PIC-PRASE_2; 1.
DR      PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KM      Isomerase; Rotamase; Membrane; Lipoprotein; Signal; Plasmid.
FT      SIGNAL 1 23
FT      CHAIN 1 299
FT      LIPID 24 24
FT      DOMAIN 144 236
FT      CONFLICT 46 46
SO      SEQUENCE 299 AA; 3313 MW; 3A7D35DE330ED5B CRC64;
Query Match 23.1%; Score 367; DB 1; Length 299;
Best Local Similarity 31.4%; Pred. No. 5,4e-13;
Matches 96; Conservative 59; Mismatches 129; Indels 22; Gaps 8;

```

```

QY      10 MKKK-----LLAGATILLSVATLAACSKGSEGLDISMGDVTEHQTFEYKSNPSAQO 64
Db      1 MKKKMKRLKVLASTATALL--LSGCOSNQDQVATYSGGKVTESNFKELKQSPPTKT 58
QY      65 VLNMNTIOKVEFEKQYSELDKEDVDITAEKKQYGENYORVLSQAGMTLETRKAQIRTS 124
Db      59 MLANMLIYRALNHAYGSAVSTKTVNDAYDSYKQOYGENDFALPSQNGFRRSSEKELRTN 118
QY      125 KLVELA---VKVAEELDEAYKKAFFDEYTPDTAQIIRLNNEDEKAEVLEKAKAGAD 181
Db      119 FLSEVALKRLKLVSEQL-----KAAMKTYQPKVTYQHILTSDEDTAKQVLSDL-AAAGKD 172
QY      182 FAOLAKNSTDEKTEKNGEITPDSASTVEPQVKKAFALVDGVSIVTATGTQAYSS 241
Db      173 FATLAKTDSIDTATKNGKISFELNNKTLDATPKDAAYKLK-----NGDYTOTPVKV-TD 227
QY      242 QYIYVLTKEKTESNIDYEKELKTVILTKONDSFTVOSIIGKELQANIKRYKQDAFO 301
Db      228 GVEYIKMINHPAKGT-FTSSKRALTAASYAKWSRDSIMQVISOVLKNOHVITIKDKDLA 286
QY      302 NIFQY 307
Db      287 DALDSY 292

```

RESULT 3

```

ID      PRSA_BACSU STANDARD; PRT; 292 AA.
AC      P24327;
DT      01-MAR-1992 (Rel. 21, Created)
DT      01-MAR-1992 (Rel. 21, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Protein export protein prsa precursor (Ec 5.2.2.18).
GN      PRSA.
OS      Bacillus subtilis.
OX      Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
RN      NCBI_TaxID=1423;
RN      [1]
RP      SEQUENCE FROM N.A.

```



```

RP CONCEPTUAL TRANSLATION.
RA Bairoch A.:
CC Unpublished observations (May-1998).
CC -1- FUNCTION: PIPIASE ACCELERATE THE FOLDING OF PROTEINS. SEEMS TO BE
CC INVOLVED IN THE FOLDING OF OUTER MEMBRANE PROTEINS (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Peptidylproline (omega-180) - peptidylproline
CC (omega-0).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE PPIC/PARVULIN FAMILY OF ROTAMASES.
CC STRONG, TO E.COLI PID.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION. A FRAMESHIFT WAS
CC INTRODUCED IN POSITION 43 TO PRODUCE THIS ORF.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32781; AAC22665.1; ALT_FRAME.
DR HSSP: O9Y237; 1E03.
DR TIGR: H11004; -.
DR InterPro: IPR000297; Rotamase.
DR Pfam: PF00639; Rotamase; 1.
DR PROSITE: PS01096; PPIC_PIPIASE_1; 1.
DR PROSITE: PS50198; PPIC_PIPIASE_2; 1.
KM Isomerase; Rotamase; Transmembrane; Periplasmic; Complete proteome.
FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 17 37 POTENTIAL.
FT DOMAIN 38 622 PERIPLASMIC (POTENTIAL).
FT DOMAIN 270 356 PPIC.
SQ SEQUENCE 622 AA; 69590 MW; EE5900A5381C4B4 CRC64;

Query Match 10.8%; Score 171; DB 1; Length 622;
Best Local Similarity 21.9%; Pred. No. 0.026;
Matches 92; Conservative 68; Mismatches 118; Indels 142; Gaps 19;

13 KLAGAITLLVATLAACSG--SEGADLISMKGDVTEHOF--YDQ----- 55
16 KFLIG---LIASFVIGMSGYLFSSNDYAAKAVNGEVIISQDFLNRVQEFETRAQRG 72
56 ---YKSNPSAQ-----QVLLNMTIQKVEKQVGSLEDKEDVDITAE----- 98
73 EAFVAQSDSPFVYTAIRQNIIVNLMDIQELROYKELKGVSDMKRAIVTDPNFQVNG 132
99 YGEN--YQRYVLSQAGMT-----LETR 117
133 KEDNAVYQRILOQNLHLSGDIYASILRASLPLEQIQNGVANSFITYAQVKNASAEVFFQKR 192
118 KQGIKTSKLV-ELAVKVAEAEITD--EAYKRAF----- 148
193 LARLATLSLADEMAKQVSDDEIKITYEANKQSFVQPEQVKQYIDLSADNISRNQVTD 252
149 ----DETPPVTAQI-----IRANDEAKVELEKAKAGDPAQLAKNDSTDEKTKEN 198
253 VETIAQYQONKAKQFMTQHLAHIQFANBODAKVAYEELQ--KGANFADAVAKAKSIDKISGEN 311
199 GGEITPDSASTPEQVKKAAFAFDVDSVITANGTQVASSQYIVYVLTKEETSSNI 258
312 GGDILGVN-ENELPKAFEDAAAALQVGYSQPIINDG-----NHVITVQR--KAQSL 362
259 DDYKEKLTQVI-----LTOKQNDSTFVQS--IIGKEIQANIKYKQDAF--QNI 303
363 ENVKQAQIADLVKRSLSMESRYFSLERQASDKAFEDSDSLMTAQAQAVKQVQESDIYSRQNV 422

```

```

DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Peptidyl-prolyl cis-trans isomerase D (EC 5.2.1.8) (PIPIase D)
DE (Rotamase D).
GN PPID OR B0441 OR C0557.
OS Escherichia coli, and
OS Escherichia coli 06.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 217992;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick K.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RN SEQUENCE FROM N.A.
RP Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,
RA Duncan M., Federspiel N., Hyman R., Kaiman S., Komp C., Kurd O.,
RA Lew H., Lin D., Namath A., Oefner P., Schramm S., Davis R.W.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN-K12 / WJ110;
RA Hatada E., Ocmori H., Qiao Y., Tsuji M., Fukuda R.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RN SEQUENCE FROM N.A.
RC STRAIN-O6:HI / CPT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasco D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli."
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
RN [5]
RN CHARACTERIZATION, AND MUTAGENESIS.
RX MEDLINE=98336187; PubMed=9670013;
RA Dartigalongue C., Raina S.;
RT "A new heat-shock gene, ppid, encodes a peptidyl-prolyl isomerase
RT required for folding of outer membrane proteins in Escherichia
RT coli."
RL EMBO J. 17:3968-3980(1998).
CC -1- FUNCTION: PIPIASE ACCELERATE THE FOLDING OF PROTEINS. SEEMS TO BE
CC INVOLVED IN THE FOLDING OF OUTER MEMBRANE PROTEINS. ITS PREFERENCE
CC AT THE P1 POSITION OF THE PEPTIDE SUBSTRATE IS GLU > LEU > ALA >
CC HIS > VAL > PHE > ILE > GLY > LYS > THR.
CC -1- CATALYTIC ACTIVITY: Peptidylproline (omega-180) - peptidylproline
CC (omega-0).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
CC -1- INDUCTION: By heat shock.
CC -1- SIMILARITY: BELONGS TO THE PPIC/PARVULIN FAMILY OF ROTAMASES.
CC STRONG, TO H.INFLUENZAE PPID.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000150; AAC73544.1; -.
DR EMBL: U82664; AAB40197.1; -.
DR EMBL: D82943; BAA11645.1; -.

```



```

Db      2 KKL---SVIFLSVMSLSGIAFADKDKVAVFYKGEVKEQIMKFEKQJNLQSGETIKNF 58
QY      60 -----PSAQVNLMTQKVEKQYSELDDKEVDPTIAEKKQYGENYQVLSQAGMTLE 115
Db      59 DDEPPDQDEKLRIKYNNLLKE-----EVAKSNITSSKEFOK-----LE 99
QY      116 TRKAQIRTSKIVELAVKVAEAEITDEAVKKAPEDEYTPDTA--QI-----IRLNEDKAK 169
Db      100 NAKNOAQOBELLANYIK-----SNITDKMDEYKRNKVGNIKGEQIKVAHIILVKSQKEAN 155
QY      170 EYLEKAKAGADPAQIAKDNSTDEKTEKNGEITF---DSASTEVEQYKKAFAALDVG 226
Db      156 DIKTKL-SKSGNETKIAEELSLDKASANGVIGYIILNQPGLVPEFEKK-AFALKVNE 213
QY      227 VS-DVTATGTQAVSSQYIVKIKTKTEKSSNIDQKELKIYIILQKNDSTFVQSIIG 285
Db      214 VSTPVKTDG-----WHITIKVLEK-----KVPPIPTKEAKVTIDNITLA 252
QY      286 KE-----LQA-ANIKV 295
Db      253 AEVLKRYIADLEKANKIKI 271

```

```

RESULT 9
CBF2_CAMJE STANDARD: PRT: 273 AA.
ID CBF2_CAMJE
AC 046105;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cell binding factor 2 precursor (Major antigen pcd4a).
GN CBF2 OR PCD4 OR C10596.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteriaceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCCTC 11168;
RC MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Kelsey J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltham T., Holtroyd S.,
RA Jagers K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences."
RL Nature 403:665-668(2000)
CC -1- SIMILARITY: BELONGS TO THE PIC/PAVULIN FAMILY OF ROTAMASES.
CC STRONG, TO H.PYLORI HP0175.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X64703; CAA59175.1;
CC EMBL: AL139075; CAB75232.1;
CC PIR: S52412; S52412.
CC HSSP: Q13526; IPIN.

```

```

DR InterPro: IPR000297; Rotamase.
DR Pfam: PF00639; Rotamase. 1.
DR PROSITE: PS01096; PPI_CPIASE_1; 1.
DR PROSITE: PS01098; PPI_CPIASE_2; 1.
KW Isomerase; Rotamase; Antigen; Signal; Complete proteome.
FT CHAIN 1 21 SIGNAL.
FT CHAIN 22 273 CELL BINDING FACTOR 2.
FT DOMAIN 131 228 PPI_C.
SQ SEQUENCE 273 AA; 30518 MW; 1C014658BBD439 CRC64;

```

```

Query Match 9.2%; Score 146; DB 1; Length 273;
Best Local Similarity 24.8%; Pred. No. 0.22;
Matches 76; Conservative 47; Mismatches 101; Indels 82; Gaps 14;

```

```

QY 10 MKR-----KLGAATLLSVATLACSGSEADLSMKGDVTE---HOPY-----E 54
Db 1 MKRFSVLAATLNGVLANNAVA-----TNGSISTEVESEFPAPLRQ 48
QY 55 QVKSNPAAOVLLMTQKVEKQYSELDDKEVDPTIAEKKQYGEN---YQVLSQAG 111
Db 49 DFKTLPDNQ-----KKALIQYIMQ-----DLIDAKKQNLKRPVTKELDR- 93
QY 112 MLETRKAOIRTSKIVELAVKVAEAEITDEAVKKAPEDEYTPD-----VTAQIIRLN 164
Db 94 -----KDALVNVYQEKILNTIKIDAKKVAFAFYQDNKDYKPARVOAKHILVAT 143
QY 165 EDKAKVELE-----KKAAGADPAQIAKDNSTDEKTEKNGEITFDSASTEVEQYKKA 219
Db 144 EKAKDIINELKGLKRELDAKFSLEAKRSIDPSGKNGDGTMTW-KPEPTDA 202
QY 220 FALDVGSDVITATGTQAVSSQYIVKIKTKTEKSSNIDQK-----EKLTVIL 270
Db 203 FALK-----NGTITTVKKNFGYHVLKENSQAKGIRKDEVKQGIENGLKFEERKVI- 257
QY 271 TOKQND 276
Db 258 NQKQGD 263

```

```

RESULT 10
PPIID_BUCAP STANDARD: PRT: 621 AA.
ID PPIID_BUCAP
AC 08K987;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Peptidyl-prolyl cis-trans isomerase D (EC 5.2.1.8) (PPIase D)
DE (Rotamase D).
GN PPIID OR BUSG462.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=98794;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=22084549; PubMed=12089438;
RA Tamas I., Klasson I., Canback B., Naeslund A.K., Eriksson A.-S.,
RA Werngreen J.U., Sandstroem J.P., Moran N.A., Andersson S.G.E.,
RT Science 296:2376-2379(2002).
CC -1- FUNCTION: PPIASES ACCELERATE THE FOLDING OF PROTEINS (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Peptidylproline (omega=180) -> peptidylproline
CC (omega=0).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE PIC/PAVULIN FAMILY OF ROTAMASES.
CC STRONG, TO E.COLI PPIID.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AE014122; AAM68005.1; .

DR InterPro; IPR000297; Rotamase.

CC Pfam; PF00639; Rotamase; 1.

DR PROSITE; PS01096; PPTC_PP1ASE_1; FALSE-NEG.

DR PROSITE; PS01098; PPTC_PP1ASE_2; 1.

KW Isomerase; Rotamase; Transmembrane; Complete proteome.

FT DOMAIN 1 10 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 11 31 POTENTIAL.

FT DOMAIN 32 621 EXTRACELLULAR (POTENTIAL).

FT DOMAIN 228 355 PPTC.

SO SEQUENCE 621 AA; 74139 MW; 5B7A89B253144C7C CRC64;

Query Match 9.2%; Score 146; DB 1; Length 621;

Best Local Similarity 25.1%; Pred. No. 0.55;

Matches 66; Conservative 51; Mismatches 80; Indels 66; Gaps 15;

QY 56 VKSNPSAQOVLNMTQKVEKQYSELDKEDVDITAE-KKQYGENQVRLSQAGMTL 114

DB IKINP-----INN--OKYTEKEINNYFDQKNEFEYPEKFKISY-----IQ 238

QY 115 ETRKAQRTSKVELAVKKAVALDEAKKAFDEYTPVTAQ--IIRLNEDKAKEV 172

DB KPKKFKIQCSN-----EIKKNWTKNIDKYSNOERQYSIOTRTKNEALSTL 286

QY 173 EKAKAEADFAQAKNDSTDEKRENGEITFPDASTPEQYKKAFAALDVGVSDVIT 232

DB 287 SELK-KGEPSFKIAKEKSIDPFSSEOGNIGWTKNF-VNNEIKIANLE-KIDQISNIT- 342

QY 233 ATGTQAVSSQYIVKLT-----KTEKSSNID-----YKELKT-----VITQKQ 274

DB 343 -----KFNNEFLIKLIKLLPKYKRISEVSDLIENIKYKSLNTYKLIKDKIAITAK 397

QY 275 NDSTFVQSIIGKELQANIKVKD 297

DB 398 NINRF--DLI--LKKTNLPKE 415

RESULT 11

SURF_HAELN STANDARD; PRT; 313 AA.

AC P44721;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DE 28-FEB-2003 (Rel. 41, Last annotation update)

GN SURF OR HI0458.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

OC Pasteurellaceae; Haemophilus.

OX NCBI_TaxID=727;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-Rd / KM20 / ATCC 51907;

RX MEDLINE=95350630; PubMed=7542800;

RA Fleisichmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M., Mckenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M., Welman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D., Uterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C., Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M., Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C.,

RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd."

RL Science 269:496-512(1995).

CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).

CC -1- SIMILARITY: BELONGS TO THE PPTC/PARVULIN FAMILY OF ROTAMASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; U32728; AAC22116.1; .

DR PIR; P64069; P64069.

DR HSSP; Q9Y237; LFJD.

DR TIGR; H10458; .

DR InterPro; IPR000297; Rotamase.

DR Pfam; PF00639; Rotamase; 1.

DR PROSITE; PS01096; PPTC_PP1ASE_1; 1.

DR PROSITE; PS01098; PPTC_PP1ASE_2; 1.

KW Isomerase; Rotamase; Signal; Complete proteome.

FT SIGNAL 1 27 POTENTIAL.

FT CHAIN 28 313 SURVIVAL PROTEIN SURF HOMOLOG.

FT DOMAIN 167 267 PPTC.

SO SEQUENCE 313 AA; 34669 MW; EAE2A7C352302450 CRC64;

Query Match 8.3%; Score 132.5; DB 1; Length 313;

Best Local Similarity 22.4%; Pred. No. 1.3;

Matches 70; Conservative 62; Mismatches 110; Indels 71; Gaps 15;

QY 12 KLLAGATITLAVATLAACSKG-SEGADLSKKGIVTTHQYEOVKSN-----PSA 62

DB 4 KKFVLRSLFLLATLGCATFSMAQAEERYVATVDGIVLE-----SOVRAMGKKGROSKI 59

QY 63 QOVLNMTQKVEKQYSELDKEDVD--DTIAEKKQYGENQVRLSQAGMTLETRK 118

DB 60 KTIIDILVQKAVOES-GVKIDPREIDHIVEDTAAANGTLTQO-FLDALDYGISLNFR 117

QY 119 AQIRT-----SKVELAVKKAVALDEAKKAFDEYTPVTAQ. 158

DB 118 QOLANQMWGAVRNKAIDESIDVTRREVVALQKMLDEKSGCTQKVTGKEY--EVNRI 175

QY 159 IIRLN--NEDRAKEVLEKAKAE--GADFAQAKNDSTDEKRENGEITFPDASTEV 211

DB 176 LKTLNPLNDQAQKQAKIRSDIAGKTTFADALAKYSKDYLSGANGSGIAPPEYIA 235

QY 212 PEQVK-----KAFA--LDVQGVSD-VITANGTQAVSSQYIVKTKKTE 253

DB 236 PQFAQTVVYKSGQVTSAPKTEFGWHILEVGVGRDGLTA--EAYTQKAEERLVNTLOQ 292

QY 254 KSSNIDYKREKLK 266

DB 293 DATN--DWYKALR 303

RESULT 12

PPID_BUCAI STANDARD; PRT; 623 AA.

AC P57550;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE 15-SEP-2003 (Rel. 42, Last annotation update)

DE Peplidy1-prolyl cis-trans isomerase D (EC 5.2.1.8) (PPIase D) (Rotamase D).

GN PPID OR B0478.

OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum symbiotic bacterium).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Buchnera.

OX NCBI_TaxID=118099;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-Tokyo 1998;

RX MEDLINE=20445173; PubMed=10993077;

RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;

RT "Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS."

RL Nature 407:81-86(2000).

```

CC -1- FUNCTION: PEPTASES ACCELERATE THE FOLDING OF PROTEINS (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Peptidylproline (omega=180) = peptidylproline
CC (omega=0).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE PPIC/PARVULIN FAMILY OF ROTAMASES.
CC STRONG, TO E.COLI PPID.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP001119: BAB13175.1: -
DR InterPro: IPR000297: Rotamase.
DR Pfam: PF00659: Rotamase; 1.
DR PROSITE: PS01096: PPIC_PPIASE_1; FALSE_NEG.
DR PROSITE: PS01098: PPIC_PPIASE_2; 1.
KM Isomerase: Rotamase; Transmembrane; Complete proteome.
FT DOMAIN 1 11 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 12 32 POTENTIAL.
FT DOMAIN 33 623 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 267 357 PPIC.
SQ SEQUENCE 623 AA; 74261 MW; 3CF307543EBE01P6 CRC64;

Query Match 8.3%; Score 131.5; DB 1; Length 623;
Best Local Similarity 19.1%; Pred. No. 3.2;
Matches 65; Conservative 62; Mismatches 92; Indels 121; Gaps 13;

QY 66 LNMATQVKEKQVSELDDEKVDITAE----- 94
DB 91 VLSGLNNVLEQYAKMMNTLVNDITIKTIYNSPIFOKNKSKRYLYLTSINSTN 150
QY 95 -----EKQYQENTQVRVLSQAGWLETRKAQT---RTSKVELAVKRV----- 134
DB 151 HEYINIKKKIKNTENLHTITSKSNFLLKKEKKIKLLSOKRIKKAIVDPSIYKNI 210
QY 135 -----AEAEI-----TDEAY-----KKADETAIPDTA----- 157
DB 211 TNOEAQIYFKKNODNFIPEKFKINVELKTDNFKHCENKEIYDWIRNTQYSTKEKR 270
QY 158 --OIRLINEDKAKVELEKAKAGADPAOLAKDNSTDEKTEKENGGETTFPSAEVPEOV 215
DB 271 RYSLIYQKNNQQAISILSRHLNTPEDFSKTAQEGSTDPISSKKGDIGWISIDL-IPDEI 329
QY 216 KKAAPALDVGVSDVITATGTQAYSQYIVKLT-----KKTESSEN 257
DB 330 KHAHLN-KKNQISDVI-----PFHNEFLIVKLNETOIGQOKIYEVFDSIKQIKQKKS 382
QY 258 IDDYKEKLVIVLTQKNDSTFVOSTIGKELOANIVKD 297
DB 383 LDLYNE-LKKNISNMLKNDPKIERI---LKNNILIOE 417

RESULT 13
LMG1_HUMAN STANDARD; PRT; 1609 AA.
AC P11047;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Laminin gamma-1 chain precursor (Laminin B2 chain).
GN LAMC1 OR LAMB2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91093128; PubMed=1985895;

```

```

RA Kallunki T., Ikonen J., Chow L.T., Kallunki P., Tryggvason K.;
RT "Structure of the human laminin B2 chain gene reveals extensive
RT divergence from the laminin B1 chain gene.";
RL J. Biol. Chem. 266:221-228(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88198245; PubMed=3360804;
RA Pikkariainen T., Kallunki T., Tryggvason K.;
RT "Human laminin B2 chain. Comparison of the complete amino acid
RT sequence with the B1 chain reveals variability in sequence homology
RT between different structural domains.";
RL J. Biol. Chem. 263:6751-6758(1988).
RN [3]
RP SEQUENCE OF 1393-1609 FROM N.A.
RX MEDLINE=89169663; PubMed=3234037;
RA Pikkariainen T., Kallunki T., Eddy R.L., Byers M.G.,
RA Haley L.L., Henry W.M., Tryggvason K., Shows T.B.;
RT "Isolation of a human laminin B2 (LAMB2) cDNA clone and assignment of
RT the gene to chromosome region 1q25-->q31.";
RL Cytogenet. Cell Genet. 48:137-141(1988).
RN [4]
RP SEQUENCE OF 1282-1609 FROM N.A.
RC TISSUE=Endothelial cells;
RX MEDLINE=92216129; PubMed=1806043;
RA Santos C.L.S., Sabbaga J., Brentani R.;
RT "Differences in human laminin B2 sequences.";
RL DNA Seq. 1:275-277(1991).
CC -1- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC is thought to mediate the attachment, migration, and organization
CC of cells into tissues during embryonic development by interacting
CC with other extracellular matrix components.
CC -1- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC different polypeptide chains (alpha, beta, gamma), which are bound
CC to each other by disulfide bonds into a cross-shaped molecule
CC comprising one long and three short arms with globules at each
CC end.
CC THE GAMMA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ),
CC LAMININ-2 (MEROSIN), LAMININ-3 (S-LAMININ), LAMININ-4 (S-MEROSIN),
CC LAMININ-6 (K-LAMININ), LAMININ-7 (KS-LAMININ).
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
CC COMPONENT).
CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -1- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.
CC -1- SIMILARITY: Contains 1 laminin EGF-like domain.
CC -1- SIMILARITY: Contains 11 laminin EGF-like domains.
CC -1- SIMILARITY: Contains 1 laminin IV domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M55210: AAA59492.1: -
DR EMBL: M55217: AAA59492.1: JOINED.
DR EMBL: M55201: AAA59492.1: JOINED.
DR EMBL: M55211: AAA59492.1: JOINED.
DR EMBL: M55212: AAA59492.1: JOINED.
DR EMBL: M55213: AAA59492.1: JOINED.
DR EMBL: M55214: AAA59492.1: JOINED.
DR EMBL: M55215: AAA59492.1: JOINED.
DR EMBL: M55216: AAA59492.1: JOINED.
DR EMBL: M55192: AAA59492.1: JOINED.
DR EMBL: M55193: AAA59492.1: JOINED.
DR EMBL: M55194: AAA59492.1: JOINED.
DR EMBL: M55195: AAA59492.1: JOINED.
DR EMBL: M55196: AAA59492.1: JOINED.
DR EMBL: M55197: AAA59492.1: JOINED.
DR EMBL: M55198: AAA59492.1: JOINED.

```

FT	DISULFID	754	770		BY SIMILARITY.	(POTENTIAL).
FT	DISULFID	773	781		BY SIMILARITY.	(POTENTIAL).
FT	DISULFID	775	792		BY SIMILARITY.	(POTENTIAL).
FT	DISULFID	795	804		BY SIMILARITY.	(POTENTIAL).
FT	DISULFID	807	825		BY SIMILARITY.	(POTENTIAL).
FT	DISULFID	828	842		BY SIMILARITY.	(POTENTIAL).
FT	DISULFID	830	849		BY SIMILARITY.	(POTENTIAL).
FT	DISULFID	852	861		BY SIMILARITY.	(POTENTIAL).
FT	DISULFID	864	881		BY SIMILARITY.	(POTENTIAL).
FT	DISULFID	884	898		BY SIMILARITY.	(POTENTIAL).
FT	DISULFID	886	905		BY SIMILARITY.	(POTENTIAL).
FT	DISULFID	907	916		BY SIMILARITY.	(POTENTIAL).
FT	DISULFID	919	932		BY SIMILARITY.	(POTENTIAL).
FT	DISULFID	935	947		BY SIMILARITY.	(POTENTIAL).
FT	DISULFID	937	954		BY SIMILARITY.	(POTENTIAL).
FT	DISULFID	956	965		BY SIMILARITY.	(POTENTIAL).
FT	DISULFID	968	980		BY SIMILARITY.	(POTENTIAL).
FT	DISULFID	983	995		BY SIMILARITY.	(POTENTIAL).
FT	DISULFID	985	1001		BY SIMILARITY.	(POTENTIAL).
FT	DISULFID	1003	1012		BY SIMILARITY.	(POTENTIAL).
FT	DISULFID	1015	1028		BY SIMILARITY.	(POTENTIAL).
FT	DISULFID	1031	1031		INTERCHAIN (PROBABLE).	(POTENTIAL).
FT	DISULFID	1034	1034		INTERCHAIN (PROBABLE).	(POTENTIAL).
FT	DISULFID	1600	1600		INTERCHAIN (PROBABLE).	(POTENTIAL).
FT	CARBOHYD	60	60		N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	134	134		N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	576	576		N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	650	650		N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	1022	1022		N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	1107	1107		N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	1161	1161		N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	1175	1175		N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	1205	1205		N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	1223	1223		N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	1241	1241		N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	1380	1380		N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	1395	1395		N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	1439	1439		N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	VARIANT	458	458		/FTID=VAR_014700.	(POTENTIAL).

Query Match 8.2% Score 129.5; DB 1; Length 1609;

Best Local Similarity 20.9%; Pred. No. 12;

Matches 65; Conservative 73; Mismatches 136; Indels 37; Gaps 13.

QY	32	KGSEADLISMKG--- <td>86</td>	86
DB	1048	KLOEESLIANGCTDEENVTDDAFEDRLK---EAREVMDDLREAGVDQNDONMDBLQ	1104
QY	87	EYVDITIAEKKQYGENYOFVISQAGMLETTRKAQT-RFSKLVELAVKYAAEELTDEAK	145
DB	1105	RVNNTLSSQISRL-QINRTIEETGNLEOARAHVENTERLTLEISRLEKKVAANVS	1165
QY	146	KAFDEYTPRYTAQIIRLN-----EKAKEVLEKAKAEGADFOLAK---DNSTDE----	193
DB	1164	VTPQSTGP-----NMNTLLAEARKLARLKHQEOADDIVYAKTANDISTEAYNLLL	1211
QY	194	KTKENGGETTFD-SASTEVEPOVKRAAFALADVQ--VSDVITATGTQA--YSOYYIYK	247
DB	1217	RTLAGENOJAFLEIELNRKEYEQAKNISODLEQOARVHEAKRAGDKAVEIYAVALSP	1276
QY	248	LTKKT--EKSSNIDYEKKELKVILTOKONDSTPVOSITIGRELQOANIKVKROAQNIPT	305
DB	1277	LDSETLENANNIMKEAMENLEOLLIDOKLDYDELDRMGKKELEVKNLLEKGTQEQTAD	1334
QY	306	OYIGGDSSSS 316	
DB	1337	QLLARADAAGA 1347	

RESULT 14
 PKNX_STRCO STANDARD; PRT: 673 AA.

AC Q9XA16; 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Probable serine/threonine-protein kinase SC03848 (EC 2.7.1.37).
 GN SC03848 OR SCH69.18.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycinae; Streptomycetaceae; Streptomycetes.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Godle A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Latke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2).";
 RT Nature 417.141-147(2002).
 CC -1- CATALYTIC ACTIVITY: ATP + a protein -> ADP + a phosphoprotein.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC -1- SIMILARITY: CONTAINS 4 PASTA DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AL939118; CAB45215.1; -
 DR PIR: T36717; T36717.
 DR InterPro: IPR005543; PASTA.
 DR InterPro: IPR000719; Prot_Kinase.
 DR InterPro: IPR002290; Ser_thr_kinase.
 DR Pfam: PF03793; PASTA. 4.
 DR Pfam: PF00069; pkinase; 1.
 DR ProDom: PD000001; Prot_kinase; 1.
 DR SMART: SM00740; PASTA. 4.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_SF; 1.
 DR Hypothetical protein: Transferase; Serine/threonine-protein kinase;
 KW ATP-binding; Repeat; Complete proteome.
 FT DOMAIN 11 277 PROTEIN KINASE.
 FT DOMAIN 379 445 PASTA 1.
 FT DOMAIN 446 511 PASTA 2.
 FT DOMAIN 512 580 PASTA 3.
 FT DOMAIN 581 649 PASTA 4.
 FT NP_BIND 17 25 ATP (BY SIMILARITY).
 FT BINDING 40 40 ATP (BY SIMILARITY).
 FT ACT_SITE 138 138 BY SIMILARITY.
 SQ SEQUENCE 673 AA; 71488 MW; 42A0615E239722DE CRC64;
 Query Match 8.0%; Score 127.5; DB 1; Length 673;
 Best Local Similarity 22.1%; Pred. No. 5.7;
 Matches 70; Conservative 45; Mismatches 151; Indels 51; Gaps 11;
 QY 14 LAGATITLIS---VATLAACSGSEGADLISMKGDTTEHGFYEQVKSNSAQOVLNMT 70
 DB 354 VLAGVYLVGAILIGKYAFSGDGGPNDKVPVPAFI-----GLSKADAQQQADNID 404
 QY 71 IQKVFKEYSGELDEVDVDTIAEKKQYKGENQVRLSQAGMTLEFRKQKIRSKIVELA 130
 DB 405 LVTFEKKQ---ECEDQPKGNICAGDPKQ-GTDVKE-STVNLVSTGAPKV-----A 451

QY 131 VKKVAEELTDEAKYKAFDEYTPDYTAQIIRLNEDKAKEVLEKAKGADPAQLAKDNS 190
 DB 452 VENVIDKNI-DEAKKQI-----EDKGEVETKQTESQSDGCTILSNPQ 493
 QY 191 TDEKTRKNGEITFPDSASTE---VPEQYKAA--FALDVDSVDTATATQAVSSQY 244
 DB 494 DPKLEKSGSTVTLLEVAKAKEKATVPDYVGRFCDEAKQAVESGDLTAVCTDQPTNDPMQ 553
 QY 245 IVKLTKTRKSSNIDYKREKTLVITLTONDNSTFVQSLTGELQAAIKVDAQFQNF 304
 DB 554 VGVKIVSTTPQSSSTQVDPGSKV--TIIVGKAVERTKVEYRGTLEARQILQOOSGFTNVO 611
 QY 305 TOYIGGDDSSSSSTSN 321
 DB 612 VAQSGSDNNKVPASN 628
 RESULT 15
 USOL YEAST STANDARD; PRT; 1790 AA.
 AC P25386; 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Intracellular protein transport protein USOL.
 GN USOL OR INT1 OR YDL058W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=X2180-1A;
 RX MEDLINE=91183402; PubMed=2010462;
 RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
 RA Yamasaki M.;
 RT "A cytoskeleton-related gene, usol, is required for intracellular
 RT protein transport in Saccharomyces cerevisiae.";
 RL J. Cell Biol. 113:245-260(1991).
 RN [2]
 RP SEQUENCE OF 782-1790 FROM N.A.
 RA Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
 RA Kendrick K.E.;
 RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-8 FROM N.A.
 RA Bai Y., Symington L.S.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
 CC COMPLEX.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
 CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
 CC ER AND THE GOLGI COMPLEX.
 CC -1- DOMAIN: THE RODLIKE TAIL. SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
 CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
 CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
 CC -1- SIMILARITY: BELONGS TO THE VDP/USOL/YBL047C FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X54378; CAA38253.1; -
 DR EMBL: L03188; AAB00143.1; -
 DR EMBL: U53668; AAB66559.1; -
 DR SGD: S0002216; USOL.
 DR InterPro: IPR002017; Spectrin.
 DR InterPro: IPR006955; Usol_p115_C.

DR InterPro: IPR006953; Usol_p115_head.
DR Pfam: PF04871; Usol_p115_C; 1.
DR Pfam: PF04869; Usol_p115_head; 1.
KM Transport: Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
FT DOMAIN 1 724 GLOBULAR HEAD.
FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
FT CONFLICT 847 847 G -> E (IN REF. 2).
FT CONFLICT 924 924 E -> K (IN REF. 2).
FT CONFLICT 1253 1253 V -> I (IN REF. 2).
FT CONFLICT 1319 1319 I -> V (IN REF. 2).
FT CONFLICT 1461 1461 N -> S (IN REF. 2).
FT CONFLICT 1581 1581 G -> S (IN REF. 2).
FT CONFLICT 1600 1600 I -> V (IN REF. 2).
FT CONFLICT 1661 1661 R -> S (IN REF. 2).
FT CONFLICT 1772 1772 D -> DEDDED (IN REF. 2).
SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;

Query Match 8.0%; Score 127.5; DB 1; Length 1790;
Best Local Similarity 22.8%; Pred. No. 17;
Matches 69; Conservative 59; Mismatches 103; Indels 71; Gaps 15;

QY 49 EHOFEQVK-----SNPSAQVILNMTIQVFEKQYG-----SELDKEVD--D 90
DB 1172 EROYNEIISQLNDEITSTQOE--NESIKRKNDELGEVAKMSTSEOSNLKKSEIDALN 1229
QY 91 TIAEKKQYGENYQRYVLSQAGMTLETRKAOI-----RTSKIVELAVK----- 132
DB 1230 LQTKELKKNENETNEASLESIKSVESEYIKIKILODCNFKKEVSELEDKLLKASEDKNS 1289
QY 133 KVAEALTPDAYKKADEYTPDYTAQIIRLNNEDEKAEVLEKAKAGADFAQLAKDNSTD 192
DB 1290 KYLELOKESSEKIKEELDATTLEKTIQLEKITNLSKAKE-----KSE-SELSRLKKTSSSE 1343
QY 193 EKTRENGEITTPDSASTVEPDYKKAFAFDVGVSDVITATGTQAYSQY-----YIVK 247
DB 1344 RKNAE-----QLEKLNEL--QIKNAFEKERKLLNEG--SSTTTOEYSEKINTLEDELIR 1396
QY 248 LFKKTE-KSSNIDYKKEKLTIV-----ILTKQNDSTFVOSIGKELOANIKYKDOAF 300
DB 1397 LQNNELKAKEIDNTRSELEKVSLSNDELLEEKQNTI-----KSLQDEILISYKDKIT 1448
QY 301 QN 302
DB 1449 RN 1450

Search completed: August 29, 2003, 10:34:25
Job time : 26 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2003, 10:30:34 ; Search time 99 seconds
(without alignments)
839.323 Million cell updates/sec

Title: US-10-049-473a-2
Perfect score: 1588
Sequence: 1 SNTYLKGYDMKKLLAGAIT.....IFTGYIGGGDSSSSSTSTNE 322

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp-invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organelle:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*
15: sp-rylirus:*
16: sp-bacteriap:*
17: sp-archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1541	97.0	313	16	097R51 streptococc
2	1538	96.9	313	16	08DQ24 streptococc
3	795	50.1	333	16	08CVC6 streptococc
4	738	46.3	309	16	08E602 streptococc
5	736	46.3	309	16	08E606 streptococc
6	696	43.8	351	16	092556 streptococc
7	687	43.3	351	16	08P055 streptococc
8	619	39.0	309	16	089X99 streptococc
9	614	38.7	309	16	08K553 streptococc
10	512	32.2	308	16	09CEV9 lactococcus
11	400.5	25.2	291	16	0929F4 listeria in
12	393.5	24.8	293	16	08Y557 listeria mo
13	373	23.5	299	2	09A103 lactococcus
14	363	22.9	295	2	053962 lactococcus
15	328	20.7	294	16	08Y759 listeria in
16	327.5	20.6	294	16	092BR2 listeria in

17	310	19.5	299	16	08CXK4 oceanobacil
18	286	18.0	333	16	09KDN4 bacillus ha
19	265.5	16.7	248	2	08GCH6 bacillus ce
20	214.5	13.5	306	16	08R760 thermoaer
21	212.5	13.4	320	16	09Y736 staphylococ
22	208	13.1	325	16	08CNR4 staphylococ
23	189.5	11.9	353	16	097E99 clostridium
24	187	11.8	677	16	097Y79 deinococcus
25	174.5	11.0	626	16	09CJW0 pasteurella
26	168	10.6	619	16	08DG31 vibrio vuln
27	163.5	10.3	623	16	08XE55 escherichia
28	163	10.3	282	16	08XPV2 ralstonia s
29	162	10.2	628	16	08ZC69 yersinia pe
30	158.5	10.0	623	16	08ZRB8 salmonella
31	158.5	10.0	623	16	08ZRB8 salmonella
32	158	9.9	619	16	09KQ10 vibrio chol
33	155.5	9.8	281	16	08XUK7 ralstonia s
34	155.5	9.8	336	16	051135 borrelia bu
35	155	9.8	303	16	098G68 rhizobium m
36	154	9.7	621	16	08EG35 shewanella
37	152	9.6	82	2	08VS96 streptococc
38	152	9.6	248	16	08XNR4 clostridium
39	151.5	9.5	342	16	08XHK0 brucella me
40	151	9.5	331	16	08YJG0 rhizobium m
41	150.5	9.5	284	16	092MJO rhizobium m
42	150.5	9.5	331	16	08FYE0 brucella su
43	150.5	9.5	7210	5	09Y7G8 drosophila
44	150.5	9.5	9270	5	08MUD9 drosophila
45	148	9.3	1434	5	081492 plasmodium

ALIGNMENTS

RESULT 1
ID Q97R51 PRELIMINARY; PRT; 313 AA.
AC Q97R51;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Protease maturation protein, putative.
DE SP0981.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TIGR4;
RX MEDLINE-21357209; PubMed-11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtapple E., Khouri H., Wolf A.M., Ullrich T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Anglucci S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.,
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
DR EMBL: AE007401; ANK45102.1;
DR TIGR: SP0981;
DR InterPro: IPR000297; Rotamase.
DR Pfam: PF00639; Rotamase; 1.
DR PROSITE: PS0198; PPIC_PP1ASE_2; 1.
KW Complete proteome.
SQ SEQUENCE 313 AA; 34412 MW; 42FEF736A0ABED0 CRC64;
Query Match 97.0%; Score 1541; DB 16; Length 313;
Best Local Similarity 100.0%; Pred. No. 3.4e-75;
Matches 313; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 10 MKKLLAGATITLLSVATTLAAGSGEGADLISMKGDVITEHOFEYQVKNPSAOQVLLNM 69
DB 1 MKKLLAGATITLLSVATTLAAGSGEGADLISMKGDVITEHOFEYQVKNPSAOQVLLNM 60
QY 70 TIQKVEKQYSGELDDKEVDPTIAEKKQYGENYQVLSQAGMTLETTRKAQIRTSKLYEL 129
DB 61 TIQKVEKQYSGELDDKEVDPTIAEKKQYGENYQVLSQAGMTLETTRKAQIRTSKLYEL 120
QY 130 AVKVAEALTDKAEVKADEYTPVTAQIIRLNNEKRAKVELEKAKAGADFAQIADKN 189
DB 121 AVKVAEALTDKAEVKADEYTPVTAQIIRLNNEKRAKVELEKAKAGADFAQIADKN 180
QY 190 STDEKTKENGGEITFDASTVEPEQVKKAFALDVGVSDVITATGTAQYSSQYIYVLT 249
DB 181 STDEKTKENGGEITFDASTVEPEQVKKAFALDVGVSDVITATGTAQYSSQYIYVLT 240
QY 250 KTEKSSNIDYKKEKTKVILTKONDSFVOSIIGKELOANIKVKDAPONIFTOYIG 309
DB 241 KTEKSSNIDYKKEKTKVILTKONDSFVOSIIGKELOANIKVKDAPONIFTOYIG 300
QY 310 GGDSSSSSSSTSN 322
DB 301 GGDSSSSSSSTSN 313

```

```

RESULT 2
Q8DQ24 PRELIMINARY: PRT; 313 AA.
AC 08DQ24:
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Proteinase maturation protein.
GN PPMa OR SPR0884.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_Taxid=171101;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA Dehoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Letkowitz E.J., Lu J., Matsushima P.,
RA McAlister S.M., McInerney M., Mcleaster K., Mundy C.W., Niclas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rocky P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.I.,
RA Glas J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6."
RT J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AE008462; AAK99688.1;
KW Complete proteome.
SQ SEQUENCE 313 AA; 34439 MW; 44063A36ABF9BED0 CRC64;

```

Query Match 96.9%; Score 1538; DB 16; Length 313;
 Best Local Similarity 99.7%; Pred. No. 5e-75;
 Matches 312; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

QY 10 MKKLLAGATITLLSVATTLAAGSGEGADLISMKGDVITEHOFEYQVKNPSAOQVLLNM 69
DB 1 MKKLLAGATITLLSVATTLAAGSGEGADLISMKGDVITEHOFEYQVKNPSAOQVLLNM 60
QY 70 TIQKVEKQYSGELDDKEVDPTIAEKKQYGENYQVLSQAGMTLETTRKAQIRTSKLYEL 129
DB 61 TIQKVEKQYSGELDDKEVDPTIAEKKQYGENYQVLSQAGMTLETTRKAQIRTSKLYEL 120
QY 130 AVKVAEALTDKAEVKADEYTPVTAQIIRLNNEKRAKVELEKAKAGADFAQIADKN 189
DB 121 AVKVAEALTDKAEVKADEYTPVTAQIIRLNNEKRAKVELEKAKAGADFAQIADKN 180

```

```

QY 190 STDEKTKENGGEITFDASTVEPEQVKKAFALDVGVSDVITATGTAQYSSQYIYVLT 249
DB 181 STDEKTKENGGEITFDASTVEPEQVKKAFALDVGVSDVITATGTAQYSSQYIYVLT 240
QY 250 KTEKSSNIDYKKEKTKVILTKONDSFVOSIIGKELOANIKVKDAPONIFTOYIG 309
DB 241 KTEKSSNIDYKKEKTKVILTKONDSFVOSIIGKELOANIKVKDAPONIFTOYIG 300
QY 310 GGDSSSSSSSTSN 322
DB 301 GGDSSSSSSSTSN 313

```

```

RESULT 3
Q8CVC6 PRELIMINARY: PRT; 333 AA.
AC 08CVC6:
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DE Putative protease maturation protein precursor.
GN PRM OR SMU.648.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_Taxid=1309;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=UAI59 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186;
RA Ajdile D., Meshan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carlson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Fetherly J.J.;
RT "Genome sequence of Streptococcus mutans JAI59, a cariogenic dental
RT pathogen."
RT Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
RL EMBL; AE014908; AAN58382.1;
DR Protease; Complete proteome.
KW SEQUENCE 333 AA; 36092 MW; 666CA84A5639C54E CRC64;

```

Query Match 50.1%; Score 795; DB 16; Length 333;
 Best Local Similarity 50.0%; Pred. No. 3.1e-35;
 Matches 156; Conservative 66; Mismatches 84; Indels 6; Gaps 2;

```

QY 10 MKKLLA-GATITLLSVATTLAAGSGEGADLISMKGDVITEHOFEYQVKNPSAOQVLLN 68
DB 1 MKKRLTATGLVTLSTVTLAAGSKTQNSKIATMGDTITVADFYEVKNSTASQAVLS 60
QY 69 MTIQKVEKQYSGELDDKEVDPTIAEKKQYGENYQVLSQAGMTLETTRKAQIRTSKLYE 128
DB 61 LTVSKVEKQYSGELDDKEVDPTIAEKKQYGENYQVLSQAGMTLETTRKAQIRTSKLYE 120
QY 129 LAVKVAEALTDKAEVKADEYTPVTAQIIRLNNEKRAKVELEKAKAGADFAQIADKN 188
DB 121 YAVKEAKKEITDASAKSAKYDQVTAQVQLDSEDAKAKSVLEPAKADGADFAKIAKD 180
QY 189 NSTDEKTKENGGEITFDASTVEPEQVKKAFALDVGVSDVITATGTAQYSSQYIYVLT 248
DB 181 NTKGKT-----EYSPDSGSTNLPQSVLSAALNLKDGSDVIKASDSTTYAPVYIYKI 235
QY 249 KTEKSSNIDYKKEKTKVILTKONDSFVOSIIGKELOANIKVKDAPONIFTOYIG 308
DB 236 TKTKDKNAMKAYKKRKLKELIYSOKLUNDSFNNAVIGKAFKANVYIKKAPSEILSOYA 295
QY 309 GGDSSSSSSSTSN 320
DB 296 AASGSGSGSGSTT 307

```

```

RESULT 4
Q8E602 PRELIMINARY: PRT; 309 AA.
AC 08E602:

```


RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
 RA Schlievert P.M., Musser J.M.;
 RT "Genome sequence of a serotype M3 strain of group A Streptococcus:
 RT phage-encoded toxins, the high-virulence phenotype, and clone
 RT emergence.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
 DR EMBL: AE006576; AAK34209.1; -
 DR EMBL: AE014157; AAK79666.1; -
 DR InterPro: IPR000297; Rotamase.
 DR PROSITE: PS50198; PPI_C_PPIASE_2; 1.
 KW Hypothetical protein; Protease; Complete proteome.
 SQ SEQUENCE 351 AA; 38533 MM; E33A620A26EEF3A8 CRC64;

Query Match 43.8%; Score 696; DB 16; Length 351;
 Best Local Similarity 44.7%; Pred. No. 6,6e-30;
 Matches 140; Conservative 62; Mismatches 105; Indels 6; Gaps 2;

QY 13 KLGAATLLSVATLACSGSEADLISKGVITEHOFYEOVKNPSAQVLLNMTIQ 72
 DB 6 KLASVYTLASVALAACQSTNDNTNTVISMKGDTISVSDPYNETKTEVSQKMLNLVLS 65
 QY 73 KVEKQYGSLELDKVDVDTIAEKKQYGENYQVLSQAGMTLETTRKAQIRTSKLVELAVK 132
 DB 66 RVEAQYGDVSKVEKAVHKTAEQYGFSAALAOSSLTPETFRKQIRSSKLVEYAVK 125
 QY 133 KVAEELTDEAYKAFDEYTPDYTAQIIRINNEDEKAEVLEKAKGADFAQLAKNDSTD 192
 DB 126 EAAKKELTDEYKAYESYPTMAVEMLTIDNETKASVLEELKAGADFTALAKEKTY- 184
 QY 193 EKTENGELTFPDASTVEPEQYKKAFAFDVGVSDVTATGQAVSSQYIVKLTCKT 252
 DB 185 --TPEKVTYKFPDGAIVNPTDVYKKAASSLNEGSDIVLSVLDPTSYQKKFYIVKVTYKA 242
 QY 253 EKSSNIDVYKERTVILTKONDSFTFVOSIIEKLEQANIKVKQDAFQNIIFYIGGCD 312
 DB 243 EKSDQYKVKRKAIIIAEKSKDMNFQNKVIANALDKANVYIKDKAFANILIAQYANLQ 302
 QY 313 ---SSSSSTSN 322
 DB 303 KTKAASESTTSE 315

RESULT 7
 Q8P0E5 PRELIMINARY; PRT; 351 AA.
 AC 08P0E5;
 DT 01-OCT-2002 (TEMBLrel. 22, Created)
 DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)
 DE Putative protease maturation protein.
 GN SPY18.1400.
 OS Streptococcus pyogenes (serotype M18).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_Taxid-186103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MGAS8232 / Serotype M18;
 RX MEDLINE-21927593; PubMed-11917108;
 RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
 RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
 RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
 RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
 RT "Genome sequence and comparative microarray analysis of serotype M18
 RT group A Streptococcus strains associated with acute rheumatic fever
 RT outbreaks.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).
 DR EMBL: AE010060; AAL97991.1; -
 DR InterPro: IPR000297; Rotamase.
 DR PROSITE: PS50198; PPI_C_PPIASE_2; 1.
 KW Protease; Complete proteome; Hypothetical protein.
 SQ SEQUENCE 351 AA; 38533 MM; 5DB0499B1FA1BC9 CRC64;

Query Match 43.3%; Score 687; DB 16; Length 351;
 Best Local Similarity 44.1%; Pred. No. 2e-29;
 Matches 136; Conservative 63; Mismatches 106; Indels 6; Gaps 2;

QY 13 KLGAATLLSVATLACSGSEADLISKGVITEHOFYEOVKNPSAQVLLNMTIQ 72
 DB 6 KLASVYTLASVALAACQSTNDNTNTVISMKGDTISVSDPYNETKTEVSQKMLNLVLS 65
 QY 73 KVEKQYGSLELDKVDVDTIAEKKQYGENYQVLSQAGMTLETTRKAQIRTSKLVELAVK 132
 DB 66 RVEAQYGDVSKVEKAVHKTAEQYGFSAALAOSSLTPETFRKQIRSSKLVEYAVK 125
 QY 133 KVAEELTDEAYKAFDEYTPDYTAQIIRINNEDEKAEVLEKAKGADFAQLAKNDSTD 192
 DB 126 EAAKKELTDEYKAYESYPTMAVEMLTIDNETKASVLEELKAGADFTALAKEKTY- 184
 QY 193 EKTENGELTFPDASTVEPEQYKKAFAFDVGVSDVTATGQAVSSQYIVKLTCKT 252
 DB 185 --TPEKVTYKFPDGAIVNPTDVYKKAASSLNEGSDIVLSVLDPTSYQKKFYIVKVTYKA 242
 QY 253 EKSSNIDVYKERTVILTKONDSFTFVOSIIEKLEQANIKVKQDAFQNIIFYIGGCD 312
 DB 243 EKSDQYKVKRKAIIIAEKSKDMNFQNKVIANALDKANVYIKDKAFANILIAQYANLQ 302
 QY 313 ---SSSSSTSN 322
 DB 303 KTKAASESTTSE 315

RESULT 8
 Q99XT9 PRELIMINARY; PRT; 309 AA.
 AC 099XT9;
 DT 01-JUN-2001 (TEMBLrel. 17, Created)
 DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
 DT 01-OCT-2002 (TEMBLrel. 22, Last annotation update)
 DE Hypothetical protein SPY2037 (Hypothetical protein spy18_2097).
 GN SPY2037 OR SPY18_2097.
 OS Streptococcus pyogenes, and
 OS Streptococcus pyogenes (serotype M18).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_Taxid-1314, 186103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SF370 / ATCC 700294 / Serotype M1;
 RX MEDLINE-21192684; PubMed-11296296;
 RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
 RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
 RA Qian X., Jia H.G., Najaf F.Z., Ren Q., Zhu H., Song L., White J.,
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
 RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MGAS8232 / Serotype M18;
 RX MEDLINE-21927593; PubMed-11917108;
 RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
 RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
 RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
 RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
 RT "Genome sequence and comparative microarray analysis of serotype M18
 RT group A Streptococcus strains associated with acute rheumatic fever
 RT outbreaks.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).
 DR EMBL: AE006625; AAK34705.1; -
 DR EMBL: AE010112; AAL98558.1; -
 DR InterPro: IPR000297; Rotamase.
 DR PROSITE: PS50198; PPI_C_PPIASE_2; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 309 AA; 34349 MM; 38DA6E0B27EALFE8 CRC64;

Query Match 39.0%; Score 619; DB 16; Length 309;

Best Local Similarity 42.6%; Pred. No. 7.5e-26;
Matches 127; Conservative 61; Mismatches 102; Indels 8; Gaps 3;

QY 13 KLLAGATLLSVATLAACSGSGADLISMGKGVITEHQFYEQVKNPSAQOVLNMTIO 72
DB 6 KLTGVVTLATVATLTAACSSSHNNTKLVSMKGDITVSDYNTKMTLQAOKMLSLVTS 65

QY 73 KYFEKQYSELDDKEVDPTIAEKKQYGENYQVLSOAGMTLETRKAQIRTSKLVELAVK 132
DB 66 RYFEYOYANKVSDKEVEKAYKQPADYGTGSFKTVLAOSGLTPETKYKQIRLTLLVEYAVK 125

QY 133 KVAEAE-LTDEAVKKAFFDEYTPDVTAQIIRLNNEDKAKEVLEKAKAGADFAQLANDST 191
DB 126 EQAKNETISKRDYRQAYDAYTPTMTAEIMQFEKEEDAKALAEVAKGADFAIAKEKT 185

QY 192 --DEKRENGGETFPDSASTEVEPOVKKAFALDVGVSDVITATGTOAVSSQYITVKT 249
DB 186 AADKKT-----TYTFSGETTLPAEYVRAASGLKEGNSREITLADPATSRTYHIHKYT 240

QY 250 KTEKSSNIDYKEKELKTVILQKQNDSTFVOSIIGKELQANIKYKQDAFONITFOY 307
DB 241 KATKAKADMKAYOKRLKDIIVTGKLPDPFQNKVIKALDKANVKIKDKAFANILAQF 298

RESULT 9
Q8K5P3 PRELIMINARY; PRT; 309 AA.
AC O8K5P3;
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, last annotation update)
DE Putative protease maturation protein.
GN PRM2.2 OR SPYM3.1740.
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=198466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MGAS315 / Serotype M3;
RX MEDLINE-22133808; PubMed-12122206;
RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Musser J.M.;
RT "Genome sequence of a serotype M3 strain of group A Streptococcus:
RT phage-encoded toxins, the high-virulence phenotype, and clone
RT emergence."
RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
DR EMBL: AE014170; AAM80347.1; -
DR InterPro: IPR000297; Rotamase.
DR PROSITE: PS50198; PRIC_PRIASE_2; 1.
KW Protease; Complete proteome; Hypothetical protein.
SQ SEQUENCE 309 AA; 34361 MW; 38DA7FEB2B1A02E8 CRC64;

Query Match 38.7%; Score 614; DB 16; Length 309;
Best Local Similarity 42.3%; Pred. No. 1.4e-25;
Matches 126; Conservative 61; Mismatches 103; Indels 8; Gaps 3;

QY 13 KLLAGATLLSVATLAACSGSGADLISMGKGVITEHQFYEQVKNPSAQOVLNMTIO 72
DB 6 KLTGVVTLATVATLTAACSSSHNNTKLVSMKGDITVSDYNTKMTLQAOKMLSLVTS 65

QY 73 KYFEKQYSELDDKEVDPTIAEKKQYGENYQVLSOAGMTLETRKAQIRTSKLVELAVK 132
DB 66 RYFEYOYANKVSDKEVEKAYKQPADYGTGSFKTVLAOSGLTPETKYKQIRLTLLVEYAVK 125

QY 133 KVAEAE-LTDEAVKKAFFDEYTPDVTAQIIRLNNEDKAKEVLEKAKAGADFAQLANDST 189
DB 126 EQAKNETISKRDYRQAYDAYTPTMTAEIMQFEKEEDAKALAEVAKGADFAIAKEKT 185

QY 190 STDEKRENGGETFPDSASTEVEPOVKKAFALDVGVSDVITATGTOAVSSQYITVKT 249
DB 186 AADKKT-----TYTFSGETTLPAEYVRAASGLKEGNSREITLADPATSRTYHIHKYT 240

DB 186 AADKKT-----TYTFSGETTLPAEYVRAASGLKEGNSREITLADPATSRTYHIHKYT 240

QY 250 KTEKSSNIDYKEKELKTVILQKQNDSTFVOSIIGKELQANIKYKQDAFONITFOY 307
DB 241 KATKAKADMKAYOKRLKDIIVTGKLPDPFQNKVIKALDKANVKIKDKAFANILAQF 298

RESULT 10
Q9CEV9 PRELIMINARY; PRT; 308 AA.
AC Q9CEV9;
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, last annotation update)
DE Protein maturation protein.
GN PAPA OR LL1725.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-111403;
RX MEDLINE-21235186; PubMed-11337471;
RA Holcetin A., Wincker P., Manger S., Jallion O., Malarme K.,
RA Weissendach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis 111403."
RL Genome Res. 11:731-753(2001).
DR EMBL: AE006403; AAK05823.1; -
KW Complete proteome.
SQ SEQUENCE 308 AA; 33832 MW; F9E7FA832E9E0B00 CRC64;

Query Match 32.2%; Score 512; DB 16; Length 308;
Best Local Similarity 36.2%; Pred. No. 4e-20;
Matches 114; Conservative 65; Mismatches 120; Indels 16; Gaps 5;

QY 10 MKKLLAGATLL-----SVATLAACSGSGADLISMGKGVITEHQFYEQVKNPS--A 62
DB 1 MKFKKLGIVMTYTFAGALAVTLSCGSSSDSASADITTMGDTIRVSDYKQKQFSPQPT 60

QY 63 QOVLNMTIOKVEKQYSELDDKEVDPTIAEKKQYGENYQVLSOAGMTLETRKAQIR 122
DB 61 NTLQNLITFEDIPTKFGKGVTDKDVSKKYSIKDQGSQFSSALQOQGLTEASFPPYMR 120

QY 123 TSKLVELAV-KKAAEDELDEAKKAFDEYTPDVTAQIIRLNNEDKAKEVLEKAKAGAD 181
DB 121 TQMLEQALADHEIKETQYTDANLKKAMESYHPDVTAIVVSESKDAATFALDAKKDDAG 180

QY 182 FAQLANDNSTDEKRENGGETFPDSASTEVEPOVKKAFALDVGVSDVITATGTOAVSS 241
DB 181 KASFERTNME-----SKVTFNSGTSTSVPEVQRAAFRLKNGEESDVETSSSTGAT 232

QY 242 QYITVKTTEKSSNIDYKEKELKTVILQKQNDSTFVOSIIGKELQANIKYKQDAFO 301
DB 233 SYIVEMVKTSTSEKGTGMNKKYKELQNVIKTEKEQDTTFVSGVIAYKLKNNVYKESARA 292

QY 302 NITFOYIGGDDSSSS 316
DB 293 SLFSQFTQTSSSSSS 307

RESULT 11
Q929F4 PRELIMINARY; PRT; 291 AA.
AC Q929F4;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, last annotation update)
DE Hypothetical protein lin2322.
GN Listeria innocua.
OS Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;

[1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-CLIP 11262 / Serovar 6a;
 RA Glaser P., Francaul L., Buchrieser C., Rusniok C., Amend A.,
 RA Baquero F., Berche P., Bloecher H., Brandt P., Chakraborty T.,
 RA Charbit A., Cherconul F., Couve E., de Darivar A., Dehoux P.,
 RA Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
 RA Entlian K.-D., Fsihl H., Garcia-del Portillo F., Garrido P.,
 RA Gauthier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
 RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kutapat G.,
 RA Madueno E., Maitounam A., Mata Vicente J., Ng E., Nedjari H.,
 RA Nordstedt G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schlueler T., Simoes N., Tierrez A.,
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,
 RT "Comparative genomics of *Listeria species*."
 RL Science 294:849-852(2001).
 DR EMBL: AL596171; CAC97550.1; -
 DR Listlist: L1N02322; -
 DR InterPro: IPR000297; Rotamase.
 DR Pfam: PF00639; Rotamase; 1.
 DR PROSITE: PSS0198; PPIc_PPIASE_2; 1.
 KM Hypothetical protein; Complete proteome.
 SO SEQUENCE 291 AA; 32359 MW; 85B0E3E5FC5A4FAL CRC64;

Query Match 25.28; Score 400.5; DB 16; Length 291;
 Best Local Similarity 33.68; Pred. No. 3.4e-14;
 Matches 103; Conservative 76; Mismatches 111; Indels 17; Gaps 11;

10 MKKKLAGATLLSVATLACSGSEADLISMGVITHEHOFEYKSNPSAQOVLNM 69
 1 MKKKLILGLVMMALFSLAC--GGGNNVKTDSGDVTDLYLDAKKDYGSEFV-QOL 56
 QY 70 TIOKVEKQYGESELDKVDVDTIAEKKQYGENYQVLSQAGMTLETRKAQIRTKLVEL 129
 DB 57 TFEKILGDKY--KVSEDDVKKFNEKYSQYGDQFSANLQSGLEKSFKSLKYNLLYOK 114
 QY 130 AVKVAEALTDKAEYKAFEDYTPDYATQIIRLNEDKAEVLEKAKAEGADPAQLAKDN 189
 DB 115 ATE--ANADTSDKAKFYETWQPDIVSHILVADENKAKEVEKELK-DGAKFADLAKY 171
 QY 190 STDEKTEKNGEITFDSASTEPVEQYKAFALDVQVSDVITATGQAVSSQYIYIKLT 249
 DB 172 STDTATKENGQOLA-PFGSGKMDPAFEKAYALKKNG--DISAPVKTQ-YG--YHIQMD 225
 QY 250 KTEKSSNIDYKELKTVILITOKNDSTFVQSIIGKELQANIKYKDAQFQNIIFYIG 309
 DB 226 KPAKTTTFEKKDRAKAEAYLASQLTEN--MORTLKEKRYKADANVKEDKLDKADFQD 283
 QY 310 GGDSSSS 316
 DB 284 SAKSDS 290

RESULT 12
 Q8Y557 PRELIMINARY; PRT; 293 AA.
 ID 08Y557
 AC 08Y557
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Hypothetical protein lmo2219.
 GN LMO2219.
 OS *Listeria monocytogenes*.
 OC Bacteria; Firmicutes; Bacilliales; Listeriaceae; *Listeria*.
 OX NCBI_TaxID=1639;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-EGD-e / Serovar 1/2a;
 RX MEDLINE=21537279; PubMed=11679669;
 RA Glaser P., Francaul L., Buchrieser C., Rusniok C., Amend A.,
 RA Baquero F., Berche P., Bloecher H., Brandt P., Chakraborty T.,
 RA Charbit A., Chetouani F., Couve E., de Darivar A., Dehoux P.,

Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
 RA Entlian K.-D., Fsihl H., Garcia-del Portillo F., Garrido P.,
 RA Gauthier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
 RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kutapat G.,
 RA Madueno E., Maitounam A., Mata Vicente J., Ng E., Nedjari H.,
 RA Nordstedt G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schlueler T., Simoes N., Tierrez A.,
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,
 RT "Comparative genomics of *Listeria species*."
 RL Science 294:849-852(2001).
 DR EMBL: AL591982; CAD00297.1; -
 DR Listlist: LMO2219; -
 DR InterPro: IPR000297; Rotamase.
 DR Pfam: PF00639; Rotamase; 1.
 DR PROSITE: PSS0198; PPIc_PPIASE_2; 1.
 KM Hypothetical protein; Complete proteome.
 SO SEQUENCE 293 AA; 32698 MW; 172EDF44A8F7E7EB CRC64;

Query Match 24.88; Score 393.5; DB 16; Length 293;
 Best Local Similarity 32.44; Pred. No. 8.2e-14;
 Matches 100; Conservative 77; Mismatches 115; Indels 17; Gaps 11;

10 MKKKLAGATLLSVATLACSGSEADLISMGVITHEHOFEYKSNPSAQOVLNM 69
 1 MKKKLILGLVMMALFSLAC--GGGNNVKTDSGDVTDLYLDAKKDYGSEFV-QOL 56
 QY 70 TIOKVEKQYGESELDKVDVDTIAEKKQYGENYQVLSQAGMTLETRKAQIRTKLVEL 129
 DB 57 TFEKILGDKY--KVSEDDVKKFNEKYSQYGDQFSANLQSGLEKSFKSLKYNLLYOK 114
 QY 130 AVKVAEALTDKAEYKAFEDYTPDYATQIIRLNEDKAEVLEKAKAEGADPAQLAKDN 189
 DB 115 ATE--ANADTSDKAKFYETWQPDIVSHILVADENKAKEVEKELK-DGAKFADLAKY 171
 QY 190 STDEKTEKNGEITFDSASTEPVEQYKAFALDVQVSDVITATGQAVSSQYIYIKLT 249
 DB 172 STDTATKENGQOLA-PFGSGKMDPAFEKAYALKKNG--DISAPVKTQ-YG--YHIQMD 225
 QY 250 KTEKSSNIDYKELKTVILITOKNDSTFVQSIIGKELQANIKYKDAQFQNIIFYIG 309
 DB 226 KPAKTTTFEKKDRAKAEAYLASQLTEN--MORTLKEKRYKADANVKEDKLDKADFQD 283
 QY 310 GGDSSSS 318
 DB 284 SSSSDSDS 292

RESULT 13
 Q9A103 PRELIMINARY; PRT; 299 AA.
 ID 09A103
 AC 09A103
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DE 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Prtm precursor.
 GN Prtm.
 OS *Lactococcus lactis* (subsp. cremoris) (*Streptococcus cremoris*).
 OC Plasmid pHP003.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; *Lactococcus*.
 OX NCBI_TaxID=1359;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-HP;
 RX MEDLINE=21367818; PubMed=11473599;
 RA Christensson C., Pillidge C.J., Ward E.J., O'Toole P.W.,
 RT "Nucleotide sequence and characterization of the cell envelope
 proteinase plasmid in *Lactococcus lactis* subsp. cremoris HP."
 RL J. Appl. Microbiol. 91:334-343(2001).
 DR EMBL: AF247159; AAK27980.1; -
 DR InterPro: IPR000297; Rotamase.
 DR Pfam: PF00639; Rotamase; 1.
 DR PROSITE: PSS0198; PPIc_PPIASE_1; 1.
 DR PROSITE: PSS0198; PPIc_PPIASE_2; 1.

Db	119	FLSEVALKLLKLVKSEQL-----KAAMKTYQPKYTVQHIILTSDBDFAKQVLSL- AAGKD	172
Oy	182	FAQIAKONSDIDEKREKNGEETTPDSASTEPDYQKKAAPALDVGSDVITATQVASS 241	
Db	173	FAMLAKTSDIDTARKNGKSGISFELNNKTTDPTDFEKAAYKRL-----NGDYIQTPEVKY- TD	227
Oy	242	QYVIVKTLTKREKSSNIDDYKERKLTJVILFKONDSFFVQSIIGKELOAANIKVQDOAFQ 301	
Db	228	GVEYIKMINHPAKGT-FTSSKALTLASVYAKWSDSSIMQVIVSQVLKNQHVITIKNDLA 286	
Oy	302	NIFQYQ 307	
Db	287	DALDSY 292	
RESULT 15			
ID	Q8Y759	PRELIMINARY; PRT; 294 AA.	
AC	Q8Y759;		
DT	01-MAR-2002 (TREMBLrel. 20, Created)		
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Hypothetical protein lmo1444.		
GN	LMO1444.		
OS	Listeria monocytogenes.		
OC	Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.		
OX	NCBI_TaxID=1639;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=EGD-e / Serovar 1/2a:		
RX	MEDLINE=21537279; PubMed=11679669;		
RA	Glaser P., Frangeul L., Buchrieser C., Rusnhol C., Amend A.,		
RA	Baquero F., Berche P., Blocher H., Brandt P., Chakraborty T.,		
RA	Charbit A., Checuanani F., Couve E., de Darvar A., Deloux P.,		
RA	Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,		
RA	Entian K.-D., Fishl H., Garcia-del Portillo F., Garrido P.,		
RA	Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,		
RA	Jones L.-M., Kaerst U., Kreft J., Kunz M., Kunst F., Kurapkat G.,		
RA	Madueno E., Maitournam A., Pablo Vicente J., Ng E., Nedjari H.,		
RA	Nordstlek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,		
RA	Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,		
RA	Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,		
RT	"Comparative genomics of Listeria species."		
RL	Science 284:849-852(2001).		
DR	EMBL; AL591979; CAC95522.1; -		
DR	Listlist; LMO01444; -		
DR	InterPro: IPR000297; Rotamase.		
DR	Pfam: PF00639; Rotamase; 1.		
DR	PROSITE: PS0198; PRIC; PRISA 2; 1.		
KW	Hypothetical protein; Complete proteome.		
SO	SEQUENCE 294 AA; 32569 MW; 190CT05A0BFB31EB7 CRC64;		
Query Match 20.7%; Score 328; DB 16; Length 294;			
Best Local Similarity 30.7%; Pred. No. 2.6e-10;			
Matches 96; Conservative 71; Mismatches 120; Indels 26; Gaps 14;			
Oy	12	KLLAGAL--TLLSVATTLAAGSKSGEGADLISMKGDTTTEHQFYQVQNSFAOQVLLNM 69	
Db	5	KKMLISVIAAFL--LLAGGSSA---VTKTDAGSVTQDELLEAMRTT-YGNVVOOL 56	
Oy	70	TIOKFEKQYQSELDDEKVEDVTIAEKKQYQENYQVLSQAQMTLETFKKAQIRTSKVEL 129	
Db	57	TFKRLIEKY--TYIEKEVNMNEYKKEYOXGDSFESTLSNNLTFTSKENLELYLVQK 114	
Oy	130	AVKVAEELTDEAYKKAFFDEYTPDYTAQIIRLNNEDKAKVELEKAKEGADFAOLAKDN 189	
Db	115	ATE--ANNQDVESKILAKYKQWEPDITVRHILVDEDAFAKEIQTFLK-NGERFIDLAKEX 171	
Oy	190	STDEKTRNGGEITPDSASTEPDYQKKAAPALD-VQSVDPYITPTQAVSSOYVIYKL 248	
Db	172	STDRTATSTNGG-LDDPFGEGMDETFEKAAYALEKNDVSGIVKST-----YG--YHLQL 224	
Oy	249	TKREKSSNIDDYKERKLTJVILFKONDSFFVQSIIGKELOAANIKVQDOAFQNIETQYI 308	

Fri Sep 5 09:24:49 2003

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2003, 19:11:17; Search time 3649 Seconds

(without alignments)
10630.008 Million cell updates/sec

Title: US-10-049-473a-1

Perfect score: 966

Sequence: 1 agtaacactatctcaag.....gcagtagtacatcaacgaa 966

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapept 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

GenEmbl:*

1: gb.ba:*

2: gb.htg:*

3: gb.in:*

4: gb.om:*

5: gb.ov:*

6: gb.pac:*

7: gb.ph:*

8: gb.pl:*

9: gb.pr:*

10: gb.ro:*

11: gb.sts:*

12: gb.sy:*

13: gb.un:*

14: gb.vl:*

15: em.ba:*

16: em.fun:*

17: em.hum:*

18: em.in:*

19: em.mu:*

20: em.om:*

21: em.or:*

22: em.ov:*

23: em.pat:*

24: em.ph:*

25: em.pl:*

26: em.ro:*

27: em.sts:*

28: em.un:*

29: em.vl:*

30: em.htg.hum:*

31: em.htg.inv:*

32: em.htg.other:*

33: em.htg.mus:*

34: em.htg.pln:*

35: em.htg.tod:*

36: em.htg.mam:*

37: em.htg.vrt:*

38: em.sy:*

39: em.htgo.hum:*

40: em.htgo.mus:*

41: em.htgo.other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	966	100.0	966	AX085885	AX085885 Sequence
2	966	100.0	966	AX138385	AX138385 Sequence
3	966	100.0	11309	AR218876	AR218876 Sequence
4	966	100.0	11309	BD003788	BD003788 Polynucle
5	966	100.0	13378	AE007401	AE007401 Streptoco
6	966	100.0	34980	AX571762	AX571762 Sequence
7	961.2	99.5	11552	AE008462	AE008462 Sequence
8	939	97.2	939	AX568726	AX568726 Sequence
9	939	97.2	111135	SPNEU1906	AR120254 Sequence
10	552.4	57.2	640	AR120254	BD063263 Streptoco
11	552.4	57.2	640	BD063263	AR193967 Sequence
12	265.6	27.5	791	AR193967	AE014908 Streptoco
13	265.4	27.5	12012	AE014908	AE006576 Streptoco
14	261.6	27.1	11615	AE01457	AE01457 Streptoco
15	260	26.9	51454	AP005143	AP005143 Streptoco
16	259.4	26.9	310950	AE014228	AE014228 Streptoco
17	259.4	26.9	20639	AX607662	AX607662 Sequence
18	257.8	26.7	6641	AX602145	AX602145 Sequence
19	257.8	26.7	167050	SA6766847	AL766847 Streptoco
20	255.2	26.4	10844	AE010060	AE010060 Streptoco
21	216.2	22.4	1110	AR228534	AR228534 Sequence
22	215.6	22.3	300225	AE016949	AE016949 Enterococ
23	211.6	21.9	10029	AE010112	AE010112 Streptoco
24	211.6	21.9	12808	AE006625	AE006625 Streptoco
25	211.6	21.7	53291	AE014170	AE014170 Streptoco
26	210	21.7	323825	AP005146	AP005146 Streptoco
27	210	21.0	970	AR228535	AR228535 Sequence
28	203.2	15.6	12898	AE006403	AE006403 Lactococc
29	150.4	12.9	201688	AC142194	AC142194 Homo sapi
30	125	9.4	302050	AL935256	AL935256 Lactobact
31	91.2	8.7	232807	SPNEU1901	AL449923 Streptoco
32	84	7.9	43980	AX413014	AX413014 Sequence
33	76.6	7.9	258650	AL596171	AL596171 Listeria
34	76.6	7.9	349980	AX417047	AX417047 Sequence
35	73.8	7.6	3720	576368	576368 ORF 5' of E
36	73.8	7.6	43658	HSV3PGEN	HSV3PGEN
37	73.8	7.6	112930	HSSEND	HSSEND
38	72	7.5	295050	AL591982	AL591982 Listeria
39	72	7.5	34980	AX641671	AX641671 Sequence
40	71.8	7.4	825	AF206632	AF206632 Plasmodiu
41	71.8	7.4	1720	AX416363	AX416363 Sequence
42	70.8	7.3	110000	AL031745	AL031745 Plasmodiu
43	70.2	7.3	110000	PFMAL13_07	PFMAL13_07 Continuat
44	67.8	7.0	304708	AE017001	AE017001 Bacillus
45					

ALIGNMENTS

RESULT 1	AX085885	966 bp	DNA	linear	PAT 09-MAR-2001
LOCUS	AX085885				
DEFINITION	Sequence 1 from Patent WO0112219.				
ACCESSION	AX085885				
VERSION	AX085885.1 GI:13275796				
KEYWORDS					
SOURCE	Streptococcus pneumoniae				
ORGANISM	Streptococcus pneumoniae				
REFERENCE	1				
AUTHORS	de Groot, R. and Hermans, P.W.				
TITLE	Pneumococcal vaccines				
JOURNAL	Patent: WO 0112219-A 1 22-FEB-2001;				

FEATURES
source
Erasmus Universiteit Rotterdam (NL)
Location/Qualifiers
1..966
/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
/db_xref="taxon:1313"
28..966
/note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAC34049.1"
/db_xref="GI:13275797"
/translation="MKKILAGATLLSVATILACSKSGEADLISMGDVITEHQFY
BOVSNPSAQOVLNMTIQKVEKQYSELDDKEVDPTAEKKQYGENYQVLSQAG
MTLETAKQIRTSKLVLAELAVKVAEALIDENYKAFDEYEPEDVQKIIIRNNEDKAK
EVLERAKAGADPAOLADNSTDENKENGSEITPDSASTEPEOVKRAAFLLVDGV
SDVITATGTOASQYIYKLTKEKSENIDYKEKTKTVILTKQNDSTFVQSLIG
KELQANIKVKNOARFONIFQYIGGSSSSSTISNE"
BASE COUNT 374 a 170 c 187 g 235 t
ORIGIN

Query Match 100.0%; Score 966; DB 6; Length 966;
Best Local Similarity 100.0%; Pred. No. 2,1e-167;
Matches 966; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGTACACTTATCTCAAGAGTAGACATGAAGAAAAATTTTGGCAGTCCCATACA 60
DB 1 AGTACACTTATCTCAAGAGTAGACATGAAGAAAAATTTTGGCAGTCCCATACA 60
QY 61 CTATTATCATGATGACAACTTTAGCAGCTTGTGGAAGGGTGCAGAGGTCAACCTTATC 120
DB 61 CTATTATCATGATGACAACTTTAGCAGCTTGTGGAAGGGTGCAGAGGTCAACCTTATC 120
QY 121 AGCATGAAGGGGATGTCATTACAGACATCAATTTTATGACAGTGAAGAACACCT 180
DB 121 AGCATGAAGGGGATGTCATTACAGACATCAATTTTATGACAGTGAAGAACACCT 180
QY 121 AGCATGAAGGGGATGTCATTACAGACATCAATTTTATGACAGTGAAGAACACCT 180
DB 121 AGCATGAAGGGGATGTCATTACAGACATCAATTTTATGACAGTGAAGAACACCT 180
QY 181 TCAGCCCAACAGTCTTGTAAATATGACATCCAAAAAGTTTGGAAAAACAATATGCG 240
DB 181 TCAGCCCAACAGTCTTGTAAATATGACATCCAAAAAGTTTGGAAAAACAATATGCG 240
QY 241 TCAGAGCTTGATGAAGAGGTGATGATGATGATGATGATGATGATGATGATGATG 300
DB 241 TCAGAGCTTGATGAAGAGGTGATGATGATGATGATGATGATGATGATGATGATG 300
QY 301 GAAACCTACCAACGCTGCTTGTCAACAGCAGTATGACTCTTGAACACGTAAGCTCA 360
DB 301 GAAACCTACCAACGCTGCTTGTCAACAGCAGTATGACTCTTGAACACGTAAGCTCA 360
QY 361 ATTCGTACAGTAATTAAGTGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 420
DB 361 ATTCGTACAGTAATTAAGTGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 420
QY 421 GATGAAGCCTATGAAGAGCCTTTGATGATGATGATGATGATGATGATGATGATG 480
DB 421 GATGAAGCCTATGAAGAGCCTTTGATGATGATGATGATGATGATGATGATGATG 480
QY 481 CGTCTTAATTAAGATGAAGCCCAAGAGTTCGAAAAAGCCCAAGCAAGAGTCT 540
DB 481 CGTCTTAATTAAGATGAAGCCCAAGAGTTCGAAAAAGCCCAAGCAAGAGTCT 540
QY 541 GATTTTGTCAATTAAGCAAGATTAATCACTGATGAAGAAAAAAGAAAAATGTTGA 600
DB 541 GATTTTGTCAATTAAGCAAGATTAATCACTGATGAAGAAAAAAGAAAAATGTTGA 600
QY 601 GAAATTAACCTTGAATTCGCTTCAACAGAGTACGAGCAAGTCAAAAAAGCCCTTTC 660
DB 601 GAAATTAACCTTGAATTCGCTTCAACAGAGTACGAGCAAGTCAAAAAAGCCCTTTC 660
QY 661 GCTTTAGATGTGATGTGTTCTCATATGATGATGATGATGATGATGATGATGATG 720
DB 661 GCTTTAGATGTGATGTGTTCTCATATGATGATGATGATGATGATGATGATGATG 720
QY 720 GCTTTAGATGTGATGTGTTCTCATATGATGATGATGATGATGATGATGATGATG 720
DB 720 GCTTTAGATGTGATGTGTTCTCATATGATGATGATGATGATGATGATGATGATG 720

QY 721 AGCCATATTTACATTTGTAAACTACATAGAAAAAGAAAAATCATTAATATGATGAC 780
DB 721 AGCCATATTTACATTTGTAAACTACATAGAAAAAGAAAAATCATTAATATGATGAC 780
QY 781 TACAAAGAAAAATTAATAAAGCTTATCTTGAAGTCAAAAAAGATGATGATGATG 840
DB 781 TACAAAGAAAAATTAATAAAGCTTATCTTGAAGTCAAAAAAGATGATGATGATG 840
QY 841 CAAGCATATTCGAAAAAGATTTGCAAGCAGCCATATTCAGTTAAGGACCAAGCTTC 900
DB 841 CAAGCATATTCGAAAAAGATTTGCAAGCAGCCATATTCAGTTAAGGACCAAGCTTC 900
QY 901 CAAATATCTTTACCAATATATATCGGTGAGATTCAGGCTCAAGCAGTATGATCA 960
DB 901 CAAATATCTTTACCAATATATATCGGTGAGATTCAGGCTCAAGCAGTATGATCA 960
QY 961 AAGCAA 966
DB 961 AAGCAA 966

RESULT 2
AX138385 966 bp DNA linear PAT 30-MAY-2001
LOCUS AX138385
DEFINITION Sequence 1 from Patent EP1075841.
ACCESSION AX138385
VERSION AX138385.1 GI:14274335
KEYWORDS
SOURCE
ORGANISM
Streptococcus pneumoniae
Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 de Groot, R. and Hermans, P.W.
Pneumococcal vaccines
Patent: EP 1075841-A 1 14-FEB-2001;
JOURNAL Erasmus Universiteit Rotterdam (NL)
FEATURES
source
1..966
/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
/db_xref="taxon:1313"
/note="The S. pneumoniae nucleotides 7632-8597 on contig 33".

BASE COUNT 374 a 170 c 187 g 235 t
ORIGIN
Query Match 100.0%; Score 966; DB 6; Length 966;
Best Local Similarity 100.0%; Pred. No. 2,1e-167;
Matches 966; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGTACACTTATCTCAAGAGTAGACATGAAGAAAAATTTTGGCAGTCCCATACA 60
DB 1 AGTACACTTATCTCAAGAGTAGACATGAAGAAAAATTTTGGCAGTCCCATACA 60
QY 61 CTATTATCATGATGACAACTTTAGCAGCTTGTGGAAGGGTGCAGAGGTCAACCTTATC 120
DB 61 CTATTATCATGATGACAACTTTAGCAGCTTGTGGAAGGGTGCAGAGGTCAACCTTATC 120
QY 121 AGCATGAAGGGGATGTCATTACAGACATCAATTTTATGACAGTGAAGAACACCT 180
DB 121 AGCATGAAGGGGATGTCATTACAGACATCAATTTTATGACAGTGAAGAACACCT 180
QY 181 TCAGAGCTTGATGAAGAGGTGATGATGATGATGATGATGATGATGATGATGATG 240
DB 181 TCAGAGCTTGATGAAGAGGTGATGATGATGATGATGATGATGATGATGATGATG 240
QY 241 TCAGAGCTTGATGAAGAGGTGATGATGATGATGATGATGATGATGATGATGATG 300
DB 241 TCAGAGCTTGATGAAGAGGTGATGATGATGATGATGATGATGATGATGATGATG 300
QY 301 GAAACCTACCAACGCTGCTTGTCAACAGCAGTATGACTCTTGAACACGTAAGCTCA 360
DB 301 GAAACCTACCAACGCTGCTTGTCAACAGCAGTATGACTCTTGAACACGTAAGCTCA 360

Accession	Source	Organism	Reference	Title	Journal	Features
DB	301	GAAACACGACAAAGCGTCTGTCTGCACAAAGCAGGTATGACTCTTGAACAACGTTAAAGCTCAA	360			
OY	361	ATTGCTACAACTAAATTAATTAAGTGAATGGACACTTAAGAAGGTAGCAGAAGCTGAATTGACA	420			
DB	361	ATTGCTACAACTAAATTAATTAAGTGAATGGACACTTAAGAAGGTAGCAGAAGCTGAATTGACA	420			
OY	421	GATGAAGCGCTTAAGAAAAGCCTTTGATGAGTACACTCCAGATGTTAAGCGCTCAATCATC	480			
DB	421	GATGAAGCGCTTAAGAAAAGCCTTTGATGAGTACACTCCAGATGTTAAGCGCTCAATCATC	480			
OY	481	CGTCTTAATTAATGAAGATTAAGGCCAAGAAGTCTCGAAAAAGCCCAAGGCAAGAGTGCT	540			
DB	481	CGTCTTAATTAATGAAGATTAAGGCCAAGAAGTCTCGAAAAAGCCCAAGGCAAGAGTGCT	540			
OY	541	GATTTTGGTCAATTAAGCCAAAGATTAATTCACACTGATGAAAAAACAAAAGAAAATGGTGGGA	600			
DB	541	GATTTTGGTCAATTAAGCCAAAGATTAATTCACACTGATGAAAAAACAAAAGAAAATGGTGGGA	600			
OY	601	GAAATTAACCTTGAATTCGCTTCAACAGAGTACCTGAGCAAGTCAAAAAAGCCGCTTC	660			
DB	601	GAAATTAACCTTGAATTCGCTTCAACAGAGTACCTGAGCAAGTCAAAAAAGCCGCTTC	660			
OY	661	GCTTTAGATGTGATGATGCTGTTCTTGATGATTAACAGCACTGGCACAAGCCTACAGT	720			
DB	661	GCTTTAGATGTGATGATGCTGTTCTTGATGATTAACAGCACTGGCACAAGCCTACAGT	720			
OY	721	AGCCAAATTTTCATATGTTAAACACTACATTAAGAAAACAGAAAATATCATTAATTTGATGATAC	780			
DB	721	AGCCAAATTTTCATATGTTAAACACTACATTAAGAAAACAGAAAATATCATTAATTTGATGATAC	780			
OY	781	TACAAAGAAAAATTAATAAACCTGTTATCTTGACTCAAAAACAAAATGATTCACAATTTGTT	840			
DB	781	TACAAAGAAAAATTAATAAACCTGTTATCTTGACTCAAAAACAAAATGATTCACAATTTGTT	840			
OY	841	CAAAGCATATTCGGAAGAAATTCGACAGCCAAATATCAAGGTTAAGGCCAACGCTTC	900			
DB	841	CAAAGCATATTCGGAAGAAATTCGACAGCCAAATATCAAGGTTAAGGCCAACGCTTC	900			
OY	901	CAAAATATCTTTACCCAAATATATCGTGTGAGATTCAGCTCAAGCAGTAGTACATCA	960			
DB	901	CAAAATATCTTTACCCAAATATATCGTGTGAGATTCAGCTCAAGCAGTAGTACATCA	960			
OY	961	AACGAA 966				
DB	961	AACGAA 966				
RESULT 3						
LOCUS	AR218876	11309 bp	DNA	linear	PAT 25-SEP-2002	
DEFINITION	Sequence 108 from patent US 6420135.					
ACCESSION	AR218876					
VERSION	AR218876.1	GI:23319810				
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unclassified.					
REFERENCE	1 (bases 1 to 11309)					
AUTHORS	Kunsch,C.A., Choi,G.-H., Dillon,P.S., Rosen,C.A., Barash,S.C.,					
TITLE	Fannon,M.R. and Dougherty,B.A.					
JOURNAL	Streptococcus pneumoniae polysaccharides and sequences					
FEATURES	Patent: US 6420135-A 108 16-JUL-2002;					
SOURCE	Location/Qualifiers					
	1..11309					
BASE COUNT	3557 a 2073 c 2451 g 3228 t					
ORIGIN	/organism="unknown"					

Query Match	100.0%	Score 966;	DB 6;	Length 11309;
Best Local Similarity	100.0%;	Pred. No. 1.2e-167;		
Matches 966;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	AGTAACTTTATCTCTAAAGGAGTAAACATGAAAGAAAAATTTTGGCAGTCCCATCACA	60	

Db	7632	AGTAAACCTTATCTCAAAGAGAGTACATGAAAGAAAAATTATTGGCAGAGTGCATCCACA	7691
QY	61	CTATTATCAGTAGCAACTTATAGCAGCTTGTTCGAAGGGTCAGAGGTGCAGACCTTATC	120
Db	7692	CTATTATCAGTAGCAACTTATAGCAGCTTGTTCGAAGGGTCAGAGGTGCAGACCTTATC	7751
QY	121	AGCATGAAAGGGAGTGCATATACAGAAATCATTTTATGCAAGTGAAGAAAGCAACCCCT	180
Db	7752	AGCATGAAAGGGAGTGCATATACAGAAATCATTTTATGCAAGTGAAGAAAGCAACCCCT	7811
QY	181	TCAGCCCAACAAGTCTTGTTAAATATGACATCCAAAAAGTTTTTGAAAAAACAATATGGC	240
Db	7812	TCAGCCCAACAAGTCTTGTTAAATATGACATCCAAAAAGTTTTTGAAAAAACAATATGGC	7871
QY	241	TCAGAGCTGATGATTAAGAGGTGATGATCTATTGCCGAAGAAAAACAATATAGGC	300
Db	7872	TCAGAGCTGATGATTAAGAGGTGATGATCTATTGCCGAAGAAAAACAATATAGGC	7931
QY	301	GAAACACTACCAAGCTGTTGTCACAAGCAGAGTATGACTCTTGAACACGTAAAGCTCAA	360
Db	7932	GAAACACTACCAAGCTGTTGTCACAAGCAGAGTATGACTCTTGAACACGTAAAGCTCAA	7991
QY	361	ATTGCTACAGTAAATTAAGTTGAGTTGGCAGTTAAAGAGTACAGAAAGCTGAATTGACA	420
Db	7992	ATTGCTACAGTAAATTAAGTTGAGTTGGCAGTTAAAGAGTACAGAAAGCTGAATTGACA	8051
QY	421	GATGAGAGCCTTAAGAAAAGCCTTTGATGAGTACACTCAGATGTATACGGCTCAATATATC	480
Db	8052	GATGAGAGCCTTAAGAAAAGCCTTTGATGAGTACACTCAGATGTATACGGCTCAATATATC	8111
QY	481	CGTCTTAATTAATGAGATTAAGGCCAAGAAAGTTCTCGAAAAAGCCAAAGCAGAGTGCCT	540
Db	8112	CGTCTTAATTAATGAGATTAAGGCCAAGAAAGTTCTCGAAAAAGCCAAAGCAGAGTGCCT	8171
QY	541	GATTTTGGTCAATTAGCCAAAGATTAATCACTGATGTGAAAAACAAAAGAAAATGATGGA	600
Db	8172	GATTTTGGTCAATTAGCCAAAGATTAATCACTGATGTGAAAAACAAAAGAAAATGATGGA	8231
QY	601	GAAATTTACCTTGGATTCTGTCAACAGAAATACCTGAGCAAGTCAAAAAAGCCGCTTTC	660
Db	8232	GAAATTTACCTTGGATTCTGTCAACAGAAATACCTGAGCAAGTCAAAAAAGCCGCTTTC	8291
QY	661	GCTTTAGATGAGATGCTGTTCTGATGCTGATTAACGCAACTGGCACAAAGCCTTACAGT	720
Db	8292	GCTTTAGATGAGATGCTGTTCTGATGCTGATTAACGCAACTGGCACAAAGCCTTACAGT	8351
QY	721	AGCCAAATTTACATTTGTAACCTACATTAAGAAAACAGAAAATATCATTAATATGATGAC	780
Db	8352	AGCCAAATTTACATTTGTAACCTACATTAAGAAAACAGAAAATATCATTAATATGATGAC	8411
QY	781	TACAAAGAAAAATTA AAAAAGCTGATATCTTGAAGTCAAAAACAAATGATTCACATTTGGT	840
Db	8412	TACAAAGAAAAATTA AAAAAGCTGATATCTTGAAGTCAAAAACAAATGATTCACATTTGGT	8471
QY	841	CAAAGCATTAATCGGAAAAAGATTCGACACGCAATATCAAGGTAAAGACCAAGCCTTC	900
Db	8472	CAAAGCATTAATCGGAAAAAGATTCGACACGCAATATCAAGGTAAAGACCAAGCCTTC	8531
QY	901	CAAAATATCTTTACCCCAATATATGCGTGGTGGAGATTCMAAGCTCAAGCAGTAGTACATCA	960
Db	8532	CAAAATATCTTTACCCCAATATATGCGTGGTGGAGATTCMAAGCTCAAGCAGTAGTACATCA	8591
QY	961	AACGAA 966	
Db	8592	AACGAA 8597	

RESULT 4				
BD003788				
LOCUS	BD003788	11309 bp	DNA	linear PAT 31-JAN-2002
DEFINITION	Polynucleotide of <i>Streptococcus pneumoniae</i> and sequence.			
ACCESSION	BD003788			

VERSION	BD003788.1	GI:18631749
KEYWORDS	JP 2001501833-A/108.	
SOURCE	unidentified	
ORGANISM	unidentified	
REFERENCE	unclassified.	
ADITHORS	1 (bases 1 to 11309)	
TITLE	Kunsch,C.A., Choi,G.H., Dillon,P.J., Rosen,C.A., Bara,S.C., Fannon,M. and Dougherty,B.A.	
JOURNAL	Polynucleotide of Streptococcus pneumoniae and sequence Patent: JP 2001501833-A 108 13-FEB-2001.	
COMMENT	HUMAN GENOME SCIENCES INC	
OS	Unidentified	
PN	JP 2001501833-A/108	
PD	13-FEB-2001	
PF	30-OCT-1997 JP 1998520718	
PR	31-OCT-1996 US 60/029960	
PI	CHARLES A KUNSCH, GIL H CHOI, PATRICK J DILLON, CRAIG A ROSEN, P STEVEN C BARASH,	
PT	MICHAEL FANNON, BRIAN A DOUGHERTY	
PC	C12M15/09, A01K67/027, C07K14/315, C07K16/12, C12N1/15, C12N1/19, PC C12N1/21,	
PC	C12M5/10, C12P21/02, C12Q1/68, G06F17/30, C12M15/00, C12M5/00, PC G06F15/40	
CC	Strandedness: Double;	
CC	Topology: Linear;	
FT	Key	Location/Qualifiers
FT	source	1..11309
FEATURES	source	Location/Qualifiers
BASE COUNT	3557 a 2073 c 2451 g 3228 t	
ORIGIN		
Query Match	100.0%: Score 966; DB 6; Length 11309;	
Best Local Similarity	100.0%: Pred. No. 1,2e-167;	
Matches	966; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 AGTAACACTTATGTCGAAAGAGTAGACATGAAAGAAAAATATTGTCAGGTGCCATTCACA	60
DB	7632 AGTAACACTTATGTCGAAAGAGTAGACATGAAAGAAAAATATTGTCAGGTGCCATTCACA	7691
QY	61 CTATTATCAGAGCAACTTTAGCGAGCTGTTGCGAAAGGCTGAGAAAGGTGCAGACTTATC	120
DB	7692 CTATTATCAGAGCAACTTTAGCGAGCTGTTGCGAAAGGCTGAGAAAGGTGCAGACTTATC	7751
QY	121 AGCATGAAAGGGAGTGCATTCACAGAACATCAATTTTATGAGCAAGTAAAGCAACCT	180
DB	7752 AGCATGAAAGGGAGTGCATTCACAGAACATCAATTTTATGAGCAAGTAAAGCAACCT	7811
QY	181 TCAGCCCAACAGTCTGTGTTAAATATGACCATCCAAAAAGTTTTGAAAAACAATATGCG	240
DB	7812 TCAGCCCAACAGTCTGTGTTAAATATGACCATCCAAAAAGTTTTGAAAAACAATATGCG	7871
QY	241 TCAGAGCTTGATGATTAAGAGGTGATATCTATTGCGCGAAGAAAAACAATATGCG	300
DB	7872 TCAGAGCTTGATGATTAAGAGGTGATATCTATTGCGCGAAGAAAAACAATATGCG	7933
QY	301 GAAAGCTACAGAGTCTGTGTCACAGACAGATGACTCTTGAAACAGTAAAGCTCAA	360
DB	7932 GAAAGCTACAGAGTCTGTGTCACAGACAGATGACTCTTGAAACAGTAAAGCTCAA	7991
QY	361 ATTGCTCAAGTAAATATTGTTGAGTGGCAGTTAAGAGGTGACAGAGCTGAATTGACA	420
DB	7992 ATTGCTCAAGTAAATATTGTTGAGTGGCAGTTAAGAGGTGACAGAGCTGAATTGACA	8051
QY	421 GATGAAGCTTATGAAGAAAGCTTTGATGATGACACTCCAGATGTTAAACGGCTCAATATC	480
DB	8052 GATGAAGCTTATGAAGAAAGCTTTGATGATGACACTCCAGATGTTAAACGGCTCAATATC	8111
QY	481 CGTCTTAATATGAGATAGGCCAAGCAAGTCTTCGAAAAAGCCAGGCAAGAGGTCT	540

[illegible]

JOURNAL Submitted (29-JUN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA

FEATURES

source

/organism="Streptococcus pneumoniae TIGR4"

/mol_type="genomic DNA"

/strain="TIGR4"

/db_xref="taxon:170187"

gene

/gene="SP0967"

CDS

/note="identified by Glimmer2; putative"

/codon_start=1

/product="conserved hypothetical protein TIGR0043"

/protein_id="AAK75088.1"

/db_xref="GI:14972442"

/translation="MYTMDVETGQVSKMLQOQTELEFAQAQKLGKEDKEMAVTFTV
NERSHELNLEFRNTDRPDVDSLEKPELEIAFDEEDLLENSELAEMSEPDVIGEL
FISIDKAHEQAEXGHSEFEREMGFLAVNGFHLINGDYHTPEEAEAMFGLOEELITAT
GLTRQ"

gene

/gene="SP0968"

CDS

/note="identified by match to PFM protein family HMM
PF01219"

/codon_start=1

/product="diacylglycerol kinase"

/protein_id="AAK75089.1"

/db_xref="GI:14972443"

/translation="MDSQNRKMKNRDLISLEPALTGIFTAKKEENMKHAYTAL
VVIAGFVOYSRIEMWFLSTFLVAFELINSAIENNVDLASHYHPSMLAKNAKDM
AAGAVLVSLFAALTGALIFLRIMDLF"

gene

/gene="SP0969"

CDS

/note="identified by match to TIGR protein family HMM
TIGR00436"

/codon_start=1

/product="GTP-binding protein Era"

/protein_id="AAK75090.1"

/db_xref="GI:14972444"

/translation="WTERSGFVAILGRPNVSKSTFLNHVMOQTAIMSDKAQTTNRI
MGIYTDKEQIVFDTPGIRHKPTALGDFVESASTREVDTVLFVNPADBARQGD
DMIERLKAAPVILVYVNIIDKHPDQLSQIDFERNOMDFKEIPIPSALOGNNVSR
LVDIISLDEGFQYFSPDOITDHPERLSEMYREKYLHLTREETPSHVAVVVDSMK
RDEFTDKVIRATIMVERDSOKGIIIGGAMLKIKSGMARDEIMLGDVPLETWY
KVKKNMRKDKLADFGYNERKY"

/gene="SP0970"

/gene="SP0970"

gene

/gene="SP0970"

CDS

/note="identified by match to PFM protein family HMM
PF01149"

/codon_start=1

/product="formaldehyde-oxidizing DNA glycosylase"

/protein_id="AAK75091.1"

/db_xref="GI:14972445"

/translation="WPELPVEYVCRGIEKLITGKISSISIRIPKMKITDLEPQRE
LPSQIESMGRKGLTFLYLYDKVILSHLRMEGKFTYPPDQPEKRNHAFHFEDGG
TLVLEVVRKFGTMEILVPLDLLDYVFISSKLGPESEDFDQVQSALAKSKPKSH
LLDQTLVAGLNIYDEVLMRAOVHAPRSQTLAEATAIHQITAVLAGOAVEGSGS
TIRYTNAGFEGDSWQDFHOYDKTGQCVACGCTIIEKIQGGGRTHCPCNGORD"

/gene="SP0971"

gene

/gene="SP0971"

CDS

/gene="SP0971"

/note="identified by match to PFM protein family HMM
PF01121"

/codon_start=1

/product="kinase, putative"

/protein_id="AAK75092.1"

/db_xref="GI:14972446"

/translation="MGKIGITGIGASGKSTVFNPLRQGFQVADAVVHQLKPGG
RLFEALVHFGOEITLLENGBELNPLASLIFSNDPERMSKIOGETIRRETLAREQ
LAOTEIRFMDIPLFEDODYSDFAEFTLVYVDADOVERLAKKQDSKDEASRLAA
QWPLEKKKDLASQVLDNNGNONOLINOVHILLEGGRODDQ"

/gene="SP0972"

/gene="SP0972"

/codon_start=1

/product="multi-drug resistance efflux pump"

/protein_id="AAK75093.1"

/db_xref="GI:14972447"

/translation="WTEINMKDNRIAMFNGFLGASISLVPPMPFVEMLVGSSQ
VAFAGLAISVSAISALFSPITWGLIADKYGKPKMMTRAGLAMITMGGLAFVNRITW
LIFRLNLGVAFGEVPPNATLIASQVPEKSGALGTISTGVAGTLPETGIFGLAE
LEGIRVFLVGESEFLAAITLTCIKEDFPVAKERKAIPIKELFTSVKPYLLNLF
LTSFVIOFSAOSIGPITLALVRLDQETLGVGLVYSSMGFSMSAGVGMKLGDK
VGNHRLVVAQFYSVIYILCANASSPLQILFPLFGICGALIPGNALSKMTPK
AGISRVFAFNOVFYILGCVGVPMAAGSNAQFGYHAFYATSLCVARSCLFNLIQFRT
ILKXKET"

/gene="SP0973"

/gene="SP0973"

/codon_start=1

/product="ribosomal protein L33"

/protein_id="AAK75094.1"

/db_xref="GI:14972448"

/translation="WRVKINLKSSCGSINLYTSKNSKTHPDKLEVLKCPKERRVTL
HLEK"

/gene="SP0974"

/gene="SP0974"

/codon_start=1

/product="preprotein translocase, SecE subunit, putative"

/protein_id="AAK75095.1"

/db_xref="GI:14972449"

/translation="WYNLLTILVLSVYIYIAIFMOPTRKNQSSNVFDAQSGDLFERS
KARGFEAYMORLTGLVFFWLALATVLSR"

/gene="SP0975"

/gene="SP0975"

/codon_start=1

/product="exoribonuclease, VabB/Rub family"

/protein_id="AAK75096.1"

/db_xref="GI:14972450"

/translation="WKDKREYLDQKGVYVNDLAQALGRSSKDFRELITLSMER
KHQIRFEDGSITLTKRKHEITLKGTHAKNGGFEVSLDEEDDLFVGKNDVNAV
DGDVVEVIRKVAORNNKTAABAKIIDLLESITTVGOVILDOEKPRYAGYISKNO
KISQPIYVKKKIEGTEVLEKFDKPSKHDFVAVSVADVGGSTDGIDVLEVL
ESMDIVSEPPAAVKAESVDPABOKMSEGRDLRDSITFTIGAAGADDDAVHVK
ALKNGNLEGGHIVADSVYVTEGSAIDKEALNRAATSVYVDRVVPMLPERISNGICSL
NPQNDRLQSAIMELDKRGVYVNTITOTYVTKTSFRMTSDVNDIILAGDEKREYIK
IVSSLELMAKLHETLENKRVKAGALNFTNPAKILVDQGGPVDIVLRQGIARMLTE
SEMLANETVAHFSKIDLPFIYRIHEPRAEKQKFDVYASSGSLIYGVASISQSB
ALODIMRAVEGEPPADVLSMMLLSRMOQARYSSEHNGHYGLADYTHFTSPIRYPD

```

gene
CDS
    7377..7844
    /gene="sp0976"
    /note="identified by match to TIGR protein family HMM TIGR00086"
    /codon_start=1
    /product="Sera-binding protein"
    /protein_id="AAK75097.1"
    /db_xref="gi:14972451"
    /translation="MAKGEKVVAONKRRARDYITVDLLEAGMVLTEIKSVRAARI
    NLKDFEAVKNGEVLNHNLAPEYEGNIMNDEPERRLHLHKROIKLDEIKGTG
    MTLVPLKYIKDGYAKLLGLAKGHVDKRESIRRRONDIAIVMKAIVNR"
    7860..8720

Query Match      100.0%; Score 966; DB 1; Length 13378;
Best Local Similarity 100.0%; Pred. No. 1,2e-167;
Matches 966; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

gene
1 AGTAACTATCTCTAAAGAGATACATGAGAAAAAATTTGGCAGTCCATCACA 60
12322 AGTAACTATCTCTAAAGAGATACATGAGAAAAAATTTGGCAGTCCATCACA 12381
61 CTATTATCAGTAGCACTTTAGCAGCTTGTTCGAAAGGGTCAAGAGTGCAGACTTATC 120
12382 CTATTATCAGTAGCACTTTAGCAGCTTGTTCGAAAGGGTCAAGAGTGCAGACTTATC 12441
121 AGCATGAAAGGGGATGTCATTACAGACATCAATTTTATGACAAAGTCAACCTT 180
12442 AGCATGAAAGGGGATGTCATTACAGACATCAATTTTATGACAAAGTCAACCTT 12501
181 TCAGCCCAACAAGCTTGTAAATATGACATCCAAAAAGTTTGAAGAAACAATATGCG 240
12502 TCAGCCCAACAAGCTTGTAAATATGACATCCAAAAAGTTTGAAGAAACAATATGCG 12561
241 TCAGAGCTTGATGATTAAGAGGTTGATGATGATGATGATGATGATGATGATGATG 300
12562 TCAGAGCTTGATGATTAAGAGGTTGATGATGATGATGATGATGATGATGATGATG 12621
301 GAAACCTACCAACGCTGCTTGTCAACAGCAGTATGACTCTTGAACACGTAACCTCA 360
12622 GAAACCTACCAACGCTGCTTGTCAACAGCAGTATGACTCTTGAACACGTAACCTCA 12681
361 ATTGCTACAGTAATTTAGTTGAGTTGGCAGTTAAGAGTGAAGCTGAATTGACA 420
12682 ATTGCTACAGTAATTTAGTTGAGTTGGCAGTTAAGAGTGAAGCTGAATTGACA 12741
421 GATGAAGCTTATTAAGAAAGCCTTTGATGATGATGATGATGATGATGATGATGATG 480
12742 GATGAAGCTTATTAAGAAAGCCTTTGATGATGATGATGATGATGATGATGATGATG 12801
481 CGTCTTAATATGAAGATTAAGCCCAAGAGTTCGAAAAAGCCCAAGGCGAGAGTCT 540
12802 CGTCTTAATATGAAGATTAAGCCCAAGAGTTCGAAAAAGCCCAAGGCGAGAGTCT 12861
541 GATTTGCTCAATTAAGCAAGATTAATCACTGATGATGATGATGATGATGATGATGATG 600
12862 GATTTGCTCAATTAAGCAAGATTAATCACTGATGATGATGATGATGATGATGATGATG 12921
601 GAAATTAACCTTATGATTCCTCAACAGAGTACAGCAAGTCAAAAAAGCCGCTTTC 660
12922 GAAATTAACCTTATGATTCCTCAACAGAGTACAGCAAGTCAAAAAAGCCGCTTTC 12981
661 GCTTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720
12982 GCTTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 13041
721 AGCCAATTTATCAATTTGTAATACTCACTAAGAAAAACGAAAAATCATTAATTTGATGAC 780

```

```

Db
13042 AGCCAATTTATCAATTTGTAATACTCACTAAGAAAAACGAAAAATCATTAATTTGATGAC 13101
781 TACAAAGAAAAATTAAGAAAGCTGATTCCTGATCAAAAAACAAATGATTCACATTTGTT 840
13102 TACAAAGAAAAATTAAGAAAGCTGATTCCTGATCAAAAAACAAATGATTCACATTTGTT 13161
841 CAAGCAATTAAGCAAGAAAGATTAAGCAAGCAATTAAGCAAGCAATTAAGCAAGCAAGCTTTC 900
13162 CAAGCAATTAAGCAAGAAAGATTAAGCAAGCAATTAAGCAAGCAATTAAGCAAGCAAGCTTTC 13221
901 CAATTAATCTTATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 960
13222 CAATTAATCTTATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 13281
961 AACGAA 966
13282 AACGAA 13287

RESULT 6
AX571762 34980 bp DNA linear PAT 29-NOV-2002
LOCUS AX571762
DEFINITION Sequence 4981 from Patent WO02077021.
ACCESSION AX571762
VERSION AX571762.1 GI:26003954
KEYWORDS
SOURCE
ORGANISM
Streptococcus pneumoniae
Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 Masignani V., Tettelin H. and Fraser C.
AUTHORS Streptococcus pneumoniae proteins and nucleic acids
TITLE Patent: WO 02077021-A 4981 03-OCT-2002;
JOURNAL Chiron Spa (IT) : THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
source
1.34980
/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
/db_xref="taxon:1313"
/note="seq 4979 too long: 2,162,598 bases-replaced by
following seq:-seq 4979: from 0.000.001 to 0.349.980-seq
4980: from 0.300.001 to 0.649.980-seq 4981: from 0.600.001
to 0.949.980-seq 4982: from 0.900.001 to 1.249.980-seq
4983: from 1.200.001 to 1.549.980-seq 4984: from 1.500.001
to 1.849.980-seq 4985: from 1.800.001 to 2.149.980-seq
4986: from 2.100.001 to 2.162.598"
BASE COUNT 108592 a 61526 c 75542 g 104320 t
ORIGIN
Query Match      100.0%; Score 966; DB 6; Length 34980;
Best Local Similarity 100.0%; Pred. No. 5.5e-168;
Matches 966; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db
1 AGTAACTATCTCTAAAGAGATACATGAGAAAAAATTTGGCAGTCCATCACA 60
12322 AGTAACTATCTCTAAAGAGATACATGAGAAAAAATTTGGCAGTCCATCACA 12381
61 CTATTATCAGTAGCACTTTAGCAGCTTGTTCGAAAGGGTCAAGAGTGCAGACTTATC 120
12382 CTATTATCAGTAGCACTTTAGCAGCTTGTTCGAAAGGGTCAAGAGTGCAGACTTATC 12441
121 AGCATGAAAGGGGATGTCATTACAGACATCAATTTTATGACAAAGTCAACCTT 180
12442 AGCATGAAAGGGGATGTCATTACAGACATCAATTTTATGACAAAGTCAACCTT 12501
181 TCAGCCCAACAAGCTTGTAAATATGACATCCAAAAAGTTTGAAGAAACAATATGCG 240
12502 TCAGCCCAACAAGCTTGTAAATATGACATCCAAAAAGTTTGAAGAAACAATATGCG 12561
241 TCAGAGCTTGATGATTAAGAGGTTGATGATGATGATGATGATGATGATGATGATG 300
12562 TCAGAGCTTGATGATTAAGAGGTTGATGATGATGATGATGATGATGATGATGATG 12621
301 GAAACCTACCAACGCTGCTTGTCAACAGCAGTATGACTCTTGAACACGTAACCTCA 360
12622 GAAACCTACCAACGCTGCTTGTCAACAGCAGTATGACTCTTGAACACGTAACCTCA 12681
361 ATTGCTACAGTAATTTAGTTGAGTTGGCAGTTAAGAGTGAAGCTGAATTGACA 420
12682 ATTGCTACAGTAATTTAGTTGAGTTGGCAGTTAAGAGTGAAGCTGAATTGACA 12741
421 GATGAAGCTTATTAAGAAAGCCTTTGATGATGATGATGATGATGATGATGATGATG 480
12742 GATGAAGCTTATTAAGAAAGCCTTTGATGATGATGATGATGATGATGATGATGATG 12801
481 CGTCTTAATATGAAGATTAAGCCCAAGAGTTCGAAAAAGCCCAAGGCGAGAGTCT 540
12802 CGTCTTAATATGAAGATTAAGCCCAAGAGTTCGAAAAAGCCCAAGGCGAGAGTCT 12861
541 GATTTGCTCAATTAAGCAAGATTAATCACTGATGATGATGATGATGATGATGATGATG 600
12862 GATTTGCTCAATTAAGCAAGATTAATCACTGATGATGATGATGATGATGATGATGATG 12921
601 GAAATTAACCTTATGATTCCTCAACAGAGTACAGCAAGTCAAAAAAGCCGCTTTC 660
12922 GAAATTAACCTTATGATTCCTCAACAGAGTACAGCAAGTCAAAAAAGCCGCTTTC 12981
661 GCTTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720
12982 GCTTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 13041
721 AGCCAATTTATCAATTTGTAATACTCACTAAGAAAAACGAAAAATCATTAATTTGATGAC 780

```

Db	221013	CCAGAGCGCTGATGANTAAAGAGGTGTATGATNCTATTTGGCGGAAGAAAAACAAATATGGC	221072
OY	301	GAAGAACTACCAACGCTGCTTGTTCACACAGAGATGACTCTTGAACACAGTAAAGCTCAA	360
Db	221073	GAAGAACTACCAACGCTGCTTGTTCACACAGAGATGACTCTTGAACACAGTAAAGCTCAA	221132
OY	361	ATTGCTACACAGAAATTTAGTTGAGTTGGCAGTTAAGAGGTGACGGAACCTGAAATTGACA	420
Db	221133	ATTGCTACACAGAAATTTAGTTGAGTTGGCAGTTAAGAGGTGACGGAACCTGAAATTGACA	221192
OY	421	GATGAAGCGCTATTAAGAAAAGCCTTTGATGAGTACGACTCCAGATGTAAACGGCTCAAAATCATC	480
Db	221193	GATGAAGCGCTATTAAGAAAAGCCTTTGATGAGTACGACTCCAGATGTAAACGGCTCAAAATCATC	221252
OY	481	CGCTTATATATGAAGATTAAGGCCAAAAGAACTTCGAAAAAGCCAAAGCCAGAAAGTGGCT	540
Db	221253	CGCTTATATATGAAGATTAAGGCCAAAAGAACTTCGAAAAAGCCAAAGCCAGAAAGTGGCT	221312
OY	541	GATTTTGGCTCAATTAAGCCAAAGATTAATTCACACTGATGAAAAAACAACAAATAGTGTGA	600
Db	221313	GATTTTGGCTCAATTAAGCCAAAGATTAATTCACACTGATGAAAAAACAACAAATAGTGTGA	221372
OY	601	GAATATTACCTTTGATTCTGCTTCAACACAGATGACTGAGCAAGTCAAAAAAGCCGCTTTC	660
Db	221373	GAATATTACCTTTGATTCTGCTTCAACACAGATGACTGAGCAAGTCAAAAAAGCCGCTTTC	221432
OY	661	GCTTTAGATGAGTGGATGGTGTTCGTGATGTGATTAAGCAACAGCGGCACACAGGCTTCAGT	720
Db	221433	GCTTTAGATGAGTGGATGGTGTTCGTGATGTGATTAAGCAACAGCGGCACACAGGCTTCAGT	221492
OY	721	AGCCAAATATTAACATTTGTAAAACTCACTAAGAAAAAGAAAAATCATTAATTTGATGAC	780
Db	221493	AGCCAAATATTAACATTTGTAAAACTCACTAAGAAAAAGAAAAATCATTAATTTGATGAC	221552
OY	781	TACAAAGAAAAATTTAAAAAGCTGTATCTGACTCCAAAAACAATATCAATCAATTTGTT	840
Db	221553	TACAAAGAAAAATTTAAAAAGCTGTATCTGACTCCAAAAACAATATCAATTTGTT	221612
OY	841	CAAGACATTAATCGAAAAAGAAATTTGCAAGCAGCCAAATATCAAGGTTAAGAGCACAAGCCTTC	900
Db	221613	CAAGACATTAATCGAAAAAGAAATTTGCAAGCAGCCAAATATCAAGGTTAAGAGCACAAGCCTTC	221672
OY	901	CAAAATATCTTTACCCCAATATATTCGGTGGTGGAGATTCAAAGCTCAAGCAGTAGTACATCA	960
Db	221673	CAAAATATCTTTACCCCAATATATATTCGGTGGTGGAGATTCAAAGCTCAAGCAGTAGTACATCA	221732
OY	961	AACGAA 966	
Db	221733	AACGAA 221738	
RESULT 7	AE008462	11552 bp	DNA linear BCT 13-SEP-2001
LOCUS	AE008462		
DEFINITION	Streptococcus pneumoniae R6 section 78 of 184 of the complete genome.		
ACCESSION	AE008462	AE007317	
VERSION	AE008462.1	GI:15458482	
KEYWORDS			
SOURCE	Streptococcus pneumoniae R6		
ORGANISM	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.		
REFERENCE	1 (bases 1 to 11552)		
AUTHORS	Hoshino, J. A., Alborn, W. Jr., Arnold, J., Blaszcak, L., Burgett, S., Dehoff, B. S., Estrem, S., Filiz, L., Fu, D. J., Fuller, W., Gerlinger, C., Gilmour, R., Glass, J. S., Khoja, H., Kraft, A., Lagace, R., LeBlanc, D. J., Lee, L. N., Lefkowitz, E. J., Lu, J., Matsushima, P., McAhren, S., McHenry, M., McMaster, K., Mundy, C., Nicas, T. I., Norris, F. H., O'Gara, M., Peery, R., Robertson, G. T., Rockey, P., Sun, P. M., Winkler, M. E., Yang, R., Young-Bellido, M., Zhao, G., Zook, C., Baitz, R. H., Jaskunas, S. R., Kostek, P. R. Jr., Skatrud, P. L., and Glass, J. I.		

TITLE	JOURNAL	MEDLINE	PUBMED	REFERENCE	AUTHORS
Genome of the bacterium <i>Streptococcus pneumoniae</i> strain R6	J. Bacteriol.	183 (19),	5709-5717 (2001)		
2 (bases 1 to 11552)					
Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszcak, L., Burgett, S.,					
Dehoff, B.S., Estrom, S., Fritz, L., Fu, D.-J., Fuller, W., Geringer, C.,					
Glilmour, R., Glass, J.S., Hann, A., Khaja, H., Kraft, A., Lagace, R.,					
Leblanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P.,					
McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicas, T.I.,					
Norris, F.H., O'gara, M., Peery, R., Robertson, G.T., Rokey, P.,					
Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G.,					
Zook, C., Baltz, R.H., Jaskunas, S.R., Rostock, P.R. Jr., Skatrud, P.L.					
and Glass, J.I.					
Direct Submission					
Submitted (27-JUL-2001)	Infectious Diseases Research, Eli Lilly and				
Company, Lilly Research Labs, Indianapolis, IN 46285-0438, USA					
location/Qualifiers					
1. .11552					
/organism="Streptococcus pneumoniae R6"					
/mol_type="genomic DNA"					
/strain "R6"					
/db_xref="taxon:171101"					
82. .2436					
/gene="rnr"					
/note="synonym: spr0878"					
82. .2436					
/gene="rnr"					
/codon_start=1					
/label="table=1					
/label="spr0878					
/product="Exoribonuclease R"					
/protein_id="AAK96682.1"					
/db_xref="GI:15458483"					
/translation="MKRRIEYGLQDKKRVYNDLAALGKSSKDFELIKTILMER					
KHQRFEDGSILTEAKKKHBITLKGIFHAHKNFGYSLGEGDDLPYKNDYNAI					
DGDIPEVAVIKVADRNRKGTAAEAKI IDILHSITVVGQYLDDEKRVAGYISKNO					
KISQIPVYKRSALKLEGTETVLEIDYPSKKHDFVAVYLDVGGSTGVDGLVEVL					
ESMDIIEFPEPAVVAEASVPDAPSOQDMERDLROEITFTIDGADAKLIDAVNHK					
ALKNGNLELGVHIDAVSYVYEGSALDKELNRATSYVYDTRVYPMLEPRLSNGICSL					
NPQDRLTOSALIMEIDKHGRVYNTITQYVTKFTFMTYSPVNDILAGDEKKREYHKE					
IVSIEELMATHLELMMRVKRGALNDFDNEAKILVNDKGGPNVILYRQGIARMEIE					
SPHMANETVAHEHRSKIDLPYTRIHEHPEAKERYQKIDYASSGLATITGASISQD					
ALQDIRAVGEPTADYLSMKLRSWQQAARTSENNHGTGLADYTHFTSPIRKTFD					
LVHMRIRDYGSRKEIAHEEQVLPETATOSSNERNAIEAREVEVMAKAEVMEEVY					
GEEEDAVVSVYKFGLEVELEPNVEGLIHITNLPEYHFNERDLTLGEGSGITFRVGG					
QOQIRERADKMGTEIDFSEVPSEPFVIEKGLGROSSRSGSGSNRRSDKEDKRRKS					
GRSDKRRHSQDKKKKKKRPFEYVAKKAKHGKGGKGRRTK"					
2399. .2866					
/gene="smpb"					
/note="synonym: spr0879"					
2399. .2866					
/gene="smpb"					
/codon_start=1					
/label="table=1					
/label="spr0879					
/product="Small protein B"					
/protein_id="AAK96683.1"					
/db_xref="GI:15458484"					
/translation="MAKGECKVVAONKARKADYITVDLEAGVLTGETIKSVBARH					
NLKGQFVQKNGEWMLSVNHAPVEEGINMOEERKRLLHKQLOKEQFKGGTG					
MTYIAKVTIKDGAKLLGLAKKKHDKDKESIKRQNDINRWAKVAVNR"					
2876. .3742					
/gene="lehb"					
/note="synonym: spr0880"					
2876. .3742					
/gene="lehb"					
/codon_start=1					
/label="table=1					
/label="spr0880					
/product="Tellurite resistance"					
/protein_id="AAK96684.1"					
/db_xref="GI:15458485"					

gene
CDS
/translation="MKMEKLYAKRMPLMNKQTMPEAVQCKHNTKVGWIKVYLGKA
LKFTLEEEGEVLAEHLFEAGADNPMAOPQAMHREVAATDDVEVTELEPCYDEPK
KYNTNPVHSEVLKEMQYVKGKALDIDGCGNSLFLAODPDVTAADONGLATEILIO
STVEDIDLMFVGLYDINSASIEDEYPTVSTVYLMFLQARIPAILIIONNEKISVGS
YNLIVACADIEDYPCSVNFPPTKEGELADYIKDWEIKTINENGHLMRDNENRLO
LRFKTLAKKIK"
/gene=".4768"
/note="synonym: spr0881"
3815..4768
/gene="cola"
/note="Competence induced protein A"
/codon_start=1
/transl_table=11
/label=spr0881
/product="competence protein"
/protein_id="AAK99685.1"
/db_xref="GI:15458486"
/translation="MVARDAKRGELVNVLEDKLEQAYTCPAGGQHLROGSPVTRH
FAKSLKDCDFEERESPEHLANKESLYHMLKKEKTKVQLEYPLSELKOIADVPVNGIL
ALEVQCSPLPOKIKERSEGYRSGOYVLMLOKIMLKEHLTRLOGELFESQMGF
YVWELDKQVLRKLKLYODLRGLHVOIKESYSGSLLFILLPKKOKISHTFY
SEDKICRTIKQQLTQNLFWKKEQEAIVQGENLITIGLKEMTPQIRIPYKGFQITE
QDLTSYQHFFYTYQKNPNDOKLYPAFYQYFLKNME"
4787..6589
/gene="pepB"
/note="synonym: spr0882"
4787..6589
/gene="pepB"
/EC_number="3.4.-.-"
/codon_start=1
/transl_table=11
/label=spr0882
/product="Group B oligopeptidase"
/protein_id="AAK99686.1"
/db_xref="GI:15458487"
/translation="WLDORNETNEKDTMDLSTIYPTDAMEBAIKDLTQLEIYVAQYE
GHLDSADNLEITEFSELEMEQIEKLYVYAMKNDQTRKQOEVYAKAMTYISOL
DOAFSEPEMEISEKQYADLEQAPKLOVQYHFDLQGDHVLQREBELLAGA
GEIFGASSETFALIDNADIVFPYVDDGKEVLSHGTYTRIMESKKEVRYGAOAL
YATYBOPHETAKTLQTNVKNQYRAKRYNKSARHAAIANFVESYDNLVAAYRK
HLPLHRLERSKTLIGTSLKMYVYPLISVESFTYQEAUKKAEDALAVLIGEDYL
SRVKAPESEKIDVYENOGKRSAGTSGSSTIDNAPMLNMDNDNLPTLVHETGSHM
HSSYRETQPYVYGDYSIFLAEIASTNENITTEKLLVEVEDDATRFALINVEDGFR
GTFVROQFAEFERHAIQADONGEVLSTDFLNKLIVADNOEYVGLSKSDNPEIOYEMA
RPHFYVYVYQYSTGFAAASALAEKIVHSGOEDRDYIDLKAGKSDYPLNVARKA
GVDMKEKEDYLINDAFVFERRLNEFEALVEKLGIA"
6591..7304
/gene="spr0883"
6591..7304
/gene="spr0883"
/note="Probable methyltransferase"
/codon_start=1
/transl_table=11
/label=spr0883
/product="Conserved hypothetical protein"
/protein_id="AAK99687.1"
/db_xref="GI:15458488"
/translation="MVESYSKNANHNMRPVYKEEIVLIMROKQVTFGLKELEDFE
RKNIPITPHETVAFERFLMETMOPKNILEIGTAIGFSLMAEAPNAKITTIDRNP
EMIGFAKEFADFDSRKOTLLEGDAVDLSTLESYDFVEMDAKSKYIVLPIELK
HLEGVGVYVLDIFQGGDYAKDIMEVRRQRTIYRGLOKLPATIDNPELITVPLG
DSILMKRVNADVQLSESE"
7371..8312
/gene="pma"
/note="synonym: spr0884"
7371..8312
/gene="pma"
/codon_start=1
/transl_table=11
/label=spr0884
/product="Proteinase maturation protein"
/protein_id="AAK99688.1"

gene
CDS
/db_xref="GI:15458489"
/translation="MKKRLAGATLLSVATVLAACSGSEGADLISKRGDVTIEHOFY
EOVKNNSAOOVLNMTIOKVFKEQYSELDDEVDPTIAEKKOYGENOYRLSQAG
MTLETRKAQVIRTSKQLAVKKAVALTEPAVKKADDEYTPDYTAQIILNEDDKAK
EVLEKAAEAGDPAQLAKNDSPKTEKENGELTFPDSASPEVPOVKAFAFDVGDV
SDVTATGTOVSSQYIVKTKTEKSSNIDYKREKLVILLOKONDSITFVOSIIG
KELOANIKVKKDQAFQNIFTQYIGGSSSSSSSTSN"
complement(8529..8771)
/gene="IS861-truncation"
/note="synonym: spr0885"
complement(8529..8771)
/gene="IS861-truncation"
/codon_start=1
/transl_table=11
/label=spr0885
/product="Degenerate transposase (orf2)"
/protein_id="AAK99689.1"
/db_xref="GI:15458490"
/translation="MLEQAFKEKHENTILHSDGWOYQHDYSHRFLSKGIGQASMR
KGNQDNGRMSEFGLISEMFYEQTFKSLNLE"
complement(8771..8968)
/gene="IS861-truncation"
/note="synonym: spr0886"
complement(8771..8968)
/gene="IS861-truncation"
/codon_start=1
/transl_table=11
/label=spr0886
/product="Degenerate transposase (orf1)"
/protein_id="AAK99690.1"
/db_xref="GI:15458491"
/translation="MYMNYTPKVRKKSNEFWGVITMKLTYDDKVOIYELRKQYSLEK
LSNKGINSLSRYMIKILDR"
complement(9027..9161)
repeat_region
Query Match 99.5% Score 961.2: DB 1: Length 11552:
Best Local Similarity 99.7%: Pred No. 9.1e-167; Indels 0; Gaps 0;
Matches 963: Conservative 0; Mismatches 3; Indels 0; Gaps 0;
1 AGTAACTTATCTCAAGAGGTAGACATGAGAAAAAATTATGGCAGGTGCATACCA 60
7344 AGTAACTTATCTCAAGAGGTAGACATGAGAAAAAATTATGGCAGGTGCATACCA 7403
61 CTATTATAGTAGACACTTTAGCAGCTTTCGAAAGGTCGAAAGGTGCAGACTTATC 120
7404 CTATTATAGTAGACACTTTAGCAGCTTTCGAAAGGTCGAAAGGTGCAGACTTATC 7463
121 AGCATGAAGGGAGTGCATTACAGACATCAATTTATGAGCAAGTGAAGCAACCT 180
7464 AGCATGAAGGGAGTGCATTACAGACATCAATTTATGAGCAAGTGAAGCAACCT 7523
181 TCAGCCCAACAGCTTGTAAATATGACATCAAAAAGTTTGAAGAAACAATATGGC 240
7524 TCAGCCCAACAGCTTGTAAATATGACATCAAAAAGTTTGAAGAAACAATATGGC 7583
241 TCAGAGCTTGATGAATGAAGAGCTTGATGATCTATGGCGAAGAAAAACAATATGGC 300
7584 TCAGAGCTTGATGAATGAAGAGCTTGATGATCTATGGCGAAGAAAAACAATATGGC 7643
301 GAAACTCAACAGCTGTCTGTACACAGCAGTATGACTCTTGAACACGTAAGCTCAA 360
7644 GAAACTCAACAGCTGTCTGTACACAGCAGTATGACTCTTGAACACGTAAGCTCAA 7703
361 ATTGCTACAGTAATATGTTGAGTTGGCAGTTAAGAAAGGTGACGAAGCTGAATTGCA 420
7704 ATTGCTACAGTAATATGTTGAGTTGGCAGTTAAGAAAGGTGACGAAGCTGAATTGCA 7763
421 GATGAAGCTTAAGAAAGCTTGTATGATGATGATGATGATGATGATGATGATGATG 480
7764 GATGAAGCTTAAGAAAGCTTGTATGATGATGATGATGATGATGATGATGATGATG 7823
481 CCTCTTAATATGAAGTAAAGCCAAAGAGTTCGAAAAAGCCAAAGGAGAGTCTCT 540

```

Db      7824  CGCTTAATTAATGAGATAGAGCCAAAGAGTTCTCGAAAAAGCCAGGAGAGGTCT 7883
OY      541   GATTTTGGCTAATTAGCCAAAGATTAATCAACTGATGTAAGAAAAAGAAAAGAAATGGTGA 600
Db      7884  GATTTTGGCTAATTAGCCAAAGATTAATCAACTGATGTAAGAAAAAGAAAAGAAATGGTGA 7943
OY      601   GAAATTAACCTTGAATTCCTGCTTCAACAGAGTACTGAGCAAGTCAAAAAAGCCGCTTTC 660
Db      7944  GAAATTAACCTTGAATTCCTGCTTCAACAGAGTACTGAGCAAGTCAAAAAAGCCGCTTTC 8003
OY      661   GCTTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
Db      8004  GCTTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8063
OY      721   AGCCAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
Db      8064  AGCCAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8123
OY      781   TACAAAGAAAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
Db      8124  TACAAAGAAAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8183
OY      841   CAAAGCATTAATGAGAAAAAGATGAGCAAGCAGCAATATCAAGTTAAGGACCAAGCTTC 900
Db      8184  CAAAGCATTAATGAGAAAAAGATGAGCAAGCAGCAATATCAAGTTAAGGACCAAGCTTC 8243
OY      901   CAAATATCTTACCAATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
Db      8244  CAAATATCTTACCAATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8303
OY      961   AACGAA 966
Db      8304  AACGAA 8309

RESULT 8
LOCUS   AX568726 939 bp DNA linear PAT 29-NOV-2002
DEFINITION Sequence 1933 from Patent WO02077021.
ACCESSION AX568726
VERSION AX568726.1 GI:26002429
KEYWORDS
SOURCE  Streptococcus pneumoniae
        Streptococcus pneumoniae
        Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
        Streptococcus.
REFERENCE
  1 Masignani,V., Tettelin,H. and Fraser,C.
    Streptococcus pneumoniae proteins and nucleic acids
    Patent: WO 02077021-A 1933 03-OCT-2002;
    Chiron Spa (IT) : THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
  source          1..939
                  /organism="Streptococcus pneumoniae"
                  /mol_type="genomic DNA"
                  /db_xref="taxon:1313"
BASE COUNT      363 a 165 c 182 g 229 t
ORIGIN
Query Match      97.2%; Score 939; DB 6; Length 939;
Best Local Similarity 100.0%; Pred. No. 1.9e-162;
Matches 939; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db      121  CATCAATTTTATGAGCAAGTGAAAAAGCAACCTTCAGCCCAACAAAGCTCTGTTAAATATG 180
OY      208  ACCATCCAAAAAGCTTTTGAAGAAAAATATGGCTGAGAGCTTGATGATTAAGAGGTGAT 267
Db      181  ACCATCCAAAAAGCTTTTGAAGAAAAATATGGCTGAGAGCTTGATGATTAAGAGGTGAT 240
OY      268  GATCACTATGGCCGAAGAAAAAAACAAATATGGCGAAAACTCAACAGTGTCTTGACAA 327
Db      241  GATCACTATGGCCGAAGAAAAAAACAAATATGGCGAAAACTCAACAGTGTCTTGACAA 300
OY      328  GCAGGTATGACTCTTGAACACAGTAAGCTCAATTCGTACAAATTAATTAAGTGAATG 387
Db      301  GCAGGTATGACTCTTGAACACAGTAAGCTCAATTCGTACAAATTAATTAAGTGAATG 360
OY      388  GCAGGTATGAGTGGCAAGAGCTGAATTAAGATGAAAGCTTAATTAAGAAAGCCCTGAT 447
Db      361  GCAGGTATGAGTGGCAAGAGCTGAATTAAGATGAAAGCTTAATTAAGAAAGCCCTGAT 420
OY      448  GAGTACAGTCCAGATGTAAGCGCTCAATTCATCCGCTTAATTAATTAAGATTAAGCCAAA 507
Db      421  GAGTACAGTCCAGATGTAAGCGCTCAATTCATCCGCTTAATTAATTAAGATTAAGCCAAA 480
OY      508  GAGTTCCTGAAAAAGCCAGGAGAGAGTGTGATTTTGTCTCAATTAAGCAAGATAT 567
Db      481  GAGTTCCTGAAAAAGCCAGGAGAGAGTGTGATTTTGTCTCAATTAAGCCAAAGATAT 540
OY      568  TCAACTGATGAAAAAAGAAAAAGTGGAGAAATTAAGTGAATTAAGTGAATTAAGTGAAT 627
Db      541  TCAACTGATGAAAAAAGAAAAAGTGGAGAAATTAAGTGAATTAAGTGAATTAAGTGAAT 600
OY      628  GAGTACAGTCCAGATGTAAGCGCTCAATTCATCCGCTTAATTAATTAAGATTAAGCCAAA 687
Db      601  GAGTACAGTCCAGATGTAAGCGCTCAATTCATCCGCTTAATTAATTAAGATTAAGCCAAA 660
OY      688  GTGATTAAGCAAGCACTGGCAGACACAGCTACAGTACATTAATTAAGTGAATTAAGTGAAT 747
Db      661  GTGATTAAGCAAGCACTGGCAGACACAGCTACAGTACATTAATTAAGTGAATTAAGTGAAT 720
OY      748  AAGAAACGAAAAAATCAATCTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 807
Db      721  AAGAAACGAAAAAATCAATCTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
OY      808  TTGACTCAAAAAAAGATGATTCACATTTGTTCAAGCATTAATTAAGTGAATTAAGTGAAT 867
Db      781  TTGACTCAAAAAAAGATGATTCACATTTGTTCAAGCATTAATTAAGTGAATTAAGTGAAT 840
OY      868  GCAGCCATATATCAAGGTTAAGGACCAAGCTTCGAAATATCTTTACCAATATATCGGT 927
Db      841  GCAGCCATATATCAAGGTTAAGGACCAAGCTTCGAAATATCTTTACCAATATATCGGT 900
OY      928  GGTGAGATTAAGGCTCAAGCAGTAGTATCATCAAGCAAA 966
Db      901  GGTGAGATTAAGGCTCAAGCAGTAGTATCATCAAGCAAA 939

RESULT 9
LOCUS   SPNEU1906 11135 bp DNA linear HTG 11-JUL-2001
DEFINITION Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.
ACCESSION AL449928
VERSION AL449928.1 GI:11545153
KEYWORDS HTG; HTGS_PHASE2.
SOURCE  Streptococcus pneumoniae
        Streptococcus pneumoniae
        Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
        Streptococcus.
REFERENCE
  1 Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Humbert,Y.,
    Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
    Francesco,M., Polissi,A., Buell,G., Feger,G., Garcia,E., Peltsch,M.
    and Garcia-Bustos,J.F.
    Annotated draft genomic sequence from a Streptococcus pneumoniae
    type 19F clinical isolate

```

JOURNAL Microb. Drug Resist. 7 (2), 99-125 (2001)
 MEDLINE 21335329
 PUBMED 11442348
 2 (bases 1 to 11135)
 REFERENCE
 AUTHORS Dopezo, J., Mendoza, A., Herrero, J., Caldara, F., Polissi, A., Humbert, Y., Friedl, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de Francesco, M., Buehl, G., Feger, G., Garcia, E., Feltsch, M. and Garcia-Bustos, J.F.
 TITLE Direct Submission
 JOURNAL Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A., Severo Ochoa 2, 28760 Tres Cantos, SPAIN
 COMMENT
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 FEATURES
 source
 1..11135
 /organism="Streptococcus pneumoniae"
 /mol_type="genomic DNA"
 /serotype="19F"
 /db_xref="taxon:1313"
 /clone="G54"
 BASE COUNT 32875 a 20454 c 25072 g 32680 t 54 others
 ORIGIN
 Query Match 97.2%; Score 939; DB 2; Length 111135;
 Best Local Similarity 99.5%; Pred. No. 6.4e-163;
 Matches 942; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 Oy 1 AGTAACACTTATCTCAAGAGATAGACATGAAAGAAAAATTTTGGCAGCTGCCATCACA 60
 Db 110189 ACTAACACTTATCTCAAGAGATAGACATGAAAGAAAAATTTTGGCAGCTGCCATCACA 110248
 Oy 61 CATATATCATGTGACACTTATGACAGCTTGTGGAAGGCTGGAAGGTCAGACCTTATC 120
 Db 110249 CATATATCATGTGACACTTATGACAGCTTGTGGAAGGCTGGAAGGTCAGACCTTATC 110308
 Oy 121 AGCATGAAAGGGGATGTCATTACAGAACATCAATTTATGACCAAGTGAAGCAACCTT 180
 Db 110309 AGCATGAAAGGGGATGTCATTACAGAACATCAATTTATGACCAAGTGAAGCAACCTT 110368
 Oy 181 TCAGCCCAACAGCTTGTGTAATATGACATCCAAAAGTTTGAAGAAACAATATATGCG 240
 Db 110369 TCAGCCCAACAGCTTGTGTAATATGACATCCAAAAGTTTGAAGAAACAATATATGCG 110428
 Oy 241 TCAGAGCTTGTATGATGATGAGGTTGATGATGATGATGATGATGATGATGATGATG 300
 Db 110429 TCAGAGCTTGTATGATGATGAGGTTGATGATGATGATGATGATGATGATGATG 110488
 Oy 301 GAAAGCTACCAAGCTGCTTGTGACAGAGGATGATGATGATGATGATGATGATGATG 360
 Db 110489 GAAAGCTACCAAGCTGCTTGTGACAGAGGATGATGATGATGATGATGATGATG 110548
 Oy 361 ATTGCTACAGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
 Db 110549 ATTGCTACAGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 110608
 Oy 421 GATGAGGCTATTAAGAAAGCCCTTGTATGATGATGATGATGATGATGATGATGATG 480
 Db 110609 GATGAGGCTATTAAGAAAGCCCTTGTATGATGATGATGATGATGATGATGATGATG 110668
 Oy 481 CGTCTTAATATGAAGATTAAGGCCCAAGAGTTTTCGAAAAAGCCAGGCAAGAGTGGT 540
 Db 110669 CGTCTTAATATGAAGATTAAGGCCCAAGAGTTTTCGAAAAAGCCAGGCAAGAGTGGT 110728
 Oy 541 GATTTGGCTCAATTAAGCCCAAGATTAATCACTGATGAAAAAACAAGAAATGGTGA 600
 Db 110729 GATTTGGCTCAATTAAGCCCAAGATTAATCACTGATGAAAAAACAAGAAATGGTGA 110788
 Oy 601 GAAATTAACCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
 Db 110789 GAAATTAACCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 110848

Oy 661 GCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
 Db 110849 GCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 110908
 Oy 721 AGCCATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780
 Db 110909 AGCCATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 110968
 Oy 781 TACAAAGAAAAATTAATAAAGCTGATATCTGATGATGATGATGATGATGATGATGAT 840
 Db 110969 TACAAAGAAAAATTAATAAAGCTGATATCTGATGATGATGATGATGATGATG 111028
 Oy 841 CAAGCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
 Db 111029 CAAGCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 111088
 Oy 901 CAAATATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 947
 Db 111089 CAAATATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 111135
 RESULT 10
 AR120254
 LOCUS AR120254 640 bp DNA linear PAT 16-MAY-2001
 DEFINITION Sequence 33 from patent US 6159469.
 ACCESSION AR120254
 VERSION AR120254.1 GI:14103830
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 640)
 AUTHORS Chou, G. H., Kunsch, C. A., Barash, S. C., Dillon, P. J., Dougherty, B., Fannon, M. R. and Rosen, C. A.
 TITLE Streptococcus pneumoniae antigens and vaccines
 JOURNAL Patent: US 6159469-A 33 12-DEC-2000;
 FEATURES
 source
 1..640
 /organism="unknown"
 BASE COUNT 234 a 112 c 141 g 153 t
 ORIGIN
 Query Match 57.2%; Score 552.4; DB 6; Length 640;
 Best Local Similarity 94.7%; Pred. No. 1.7e-91;
 Matches 595; Conservative 0; Mismatches 26; Indels 7; Gaps 2;
 Oy 90 TTCGAAGGGTCAGAGGTCAGACCTTATGACATGAAAGGGGATGCTATTAAGCAACA 149
 Db 1 TTCGAAGGGTCAGAGGTCAGACCTTATGACATGAAAGGGGATGCTATTAAGCAACA 60
 Oy 150 TCAATTTTATGAGCAAGTGAAGCAACCTTACAGCCCAACAAGTCTGTTAAATATGAC 209
 Db 61 TCAATTTTATGAGCAAGTGAAGCAACCTTACAGCCCAACAAGTCTGTTAAATATGAC 120
 Oy 210 CATCCAAAAAGTTTGAAGAAACAATATGCTGACAGCTTGTATGATTAAGAGAGTGTATG 269
 Db 121 CATCCAAAAAGTTTGAAGAAACAATATGCTGACAGCTTGTATGATTAAGAGAGTGTATG 180
 Oy 270 TACTATGCGGAAGAAAAAACAATATGCGGCAAACTCAACAGCTGTTGTCACAAAC 329
 Db 181 TACTATGCGGCAAGAAAAAACAATATGCGGCAAACTCAACAGCTGTTGTCACAAAC 240
 Oy 330 AGGTATGACTTGTGAACAAGTGAAGCTCAAAATGCTATCAAGTAAATTAATGATGATG 389
 Db 241 AGGTATGACTTGTGAACAAGTGAAGCTCAAAATGCTATCAAGTAAATTAATGATGATG 300
 Oy 390 AGTTAAGAGGTAGCAAGAGGTCGAATGACACATGAAGCTTATTAAGAAAGCTTTGATGA 449
 Db 301 AGTTAAGAGGTAGCAAGAGGTCGAATGACACATGAAGCTTATTAAGAAAGCTTTGATGA 360
 Oy 450 GTACACTCAGATGTAAGGCTCAATATCATGCTGTTAATTAATGAAGATTAAGGCCAAGA 509
 Db 361 GTACACTCAGATGTAAGGCTCAATATCATGCTGTTAATTAATGAAGATTAAGGCCAAGA 420


```

QY 510 AGTTCGAAAAAGCCAGCAGAGTGTGCTATTTGCTCAATTAGCCAAAGATAATTC 569
DB 421 AGTTCTCAAAAAGCCAGCAGAGTGTGCTATTTGCTCAATTAGCCAAAGATAATTC 480
QY 570 AACTGATGAAAAACAAAAAAGAAATGTGTGAGAAATTAACCTTTGATTCTGCTTCAACAGA 629
DB 481 AACTGATGAAAAACAAAAAAGAAATGTGTGAGAAATTAACCTTTGATTCTGCTTCAACAGA 540
QY 630 AGTACCTG--AGCAGTCAAAAAAGCCGCTTTCGCTTTAGATGTGATGCTGTTCTGAT 687
DB 541 AGTACCTGAGCAAGTCCAAAAAGCCGCTTTCGCTTTAGATGTGATGCTGTTCTGAT 600
QY 688 G-----TGATTACAGCAACTGGCAGACACA 710
DB 601 GGATGTGATTACAGCAACTGGGCGACACA 628

RESULT 11
BD063263 640 bp DNA linear PAT 27-AUG-2002
LOCUS Streptococcus pneumoniae antigens and vaccines.
ACCESSION BD063263.1 GI:22608866
VERSION JP 2001505415-A/17.
KEYWORDS unclassified
SOURCE unclassified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 640)
AUTHORS Kunsch,C.A., Choi,G.H., Johnson,S.L. and Hromocky,J.A.
TITLE Streptococcus pneumoniae antigens and vaccines
JOURNAL Patent: JP 2001505415-A 17 24-APR-2001;
HUMAN GENOME SCIENCES INC
PN JP 2001505415-A/17
PD 24-APR-2001
PF 30-OCT-1997 JP 1998520667
PR 31-OCT-1996 US 60/029960
PI CHARLES A KUNSCH,GIL H CHOI,SYDNOR L JOHNSON,ALEX HROMOCKYJ PC
C12N15/31,C12N5/18,C12N1/21,C07K14/315,C12Q1/68,A61K39/09, PC
G01N33/569,
PC G01N33/68
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers.
FEATURES
source 1..640
location/Qualifiers
BASE COUNT 234 a 112 c 141 g 153 t
ORIGIN

Query Match 57.2%; Score 552.4; DB 6; Length 640;
Best Local Similarity 94.7%; Pred. No. 1.7e-91;
Matches 595; Conservative 0; Mismatches 26; Indels 7; Gaps 2;

QY 90 TTCGAAAGGTCAGAGTGTGACACCTTTCAGCATGGAAGGGATGTCATACAGAA 149
DB 1 TTCGAAAGGTCAGAGTGTGACACCTTTCAGCATGGAAGGGATGTCATACAGAA 60
QY 150 TCATATTTATGACAGTGTGAAAGCAACCTTCAGCCCAACAAGTCTTTTAAATATGAC 209
DB 61 TCATATTTATGACAGTGTGAAAGCAACCTTCAGCCCAACAAGTCTTTTAAATATGAC 120
QY 210 CATCCAAAAAGTTTGAAGAAACATATGTGCTCAGAGCTTGATGATTAAGAGGTTGATGA 269
DB 121 CATCCAAAAAGTTTGAAGAAACATATGTGCTCAGAGCTTGATGATTAAGAGGTTGATGA 180
QY 270 TGTCTATGCGAAGAAAAACAATATGCGAAAACTACCAACGTCCTTGTCACAAAGC 329
DB 181 TGTCTATGCGAAGAAAAACAATATGCGAAAACTACCAACGTCCTTGTCACAAAGC 240
QY 330 AGGTATGACTTGTGAACAAGTAAAGCTCAAAATTCGTACAGTAATTAATGATGAGTTGAC 389

```

```

DB 241 AGGTATGACTTGTGAACAAGTAAAGCTCAAAATTCGTACAAAGTAATTAATGATGAGTTGAC 300
QY 390 AGTTAAGAAAGTACAGCAAGAGCTGATTTGACATATGACGCTTTAAGAAAAGCTTTGATGA 449
DB 301 AGTTAAGAAAGTACAGCAAGAGCTGATTTGACATATGACGCTTTAAGAAAAGCTTTGATGA 360
QY 450 GTACACTCCAGATGTATACGGCTCAATCATCCGCTTTTAAATATGAAGATAAGCCCAAGA 509
DB 361 GTACACTCCAGATGTATACGGCTCAATCATCCGCTTTTAAATATGAAGATAAGCCCAAGA 420
QY 510 AGTTCGAAAAAGCCAGCAGAGTGTGCTATTTGCTCAATTAGCCAAAGATAATTC 569
DB 421 AGTTCGAAAAAGCCAGCAGAGTGTGCTATTTGCTCAATTAGCCAAAGATAATTC 480
QY 570 AACTGATGAAAAACAAAAAAGAAATGTGTGAGAAATTAACCTTTGATTCTGCTTCAACAGA 629
DB 481 AACTGATGAAAAACAAAAAAGAAATGTGTGAGAAATTAACCTTTGATTCTGCTTCAACAGA 540
QY 630 AGTACCTG--AGCAGTCAAAAAAGCCGCTTTCGCTTTAGATGTGATGCTGTTCTGAT 687
DB 541 AGTACCTGAGCAAGTCCAAAAAGCCGCTTTCGCTTTAGATGTGATGCTGTTCTGAT 600
QY 688 G-----TGATTACAGCAACTGGCAGACACA 710
DB 601 GGATGTGATTACAGCAACTGGGCGACACA 628

RESULT 12
ARI93967 791 bp DNA linear PAT 20-APR-2002
LOCUS ARI93967
DEFINITION Sequence 219 from patent US 6348328.
ACCESSION ARI93967
VERSION ARI93967.1 GI:20240559
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 791)
AUTHORS Black,M.Terence., Hodgson,J.Edward., Knowles,D.Justin,Charles.,
Nicholas,R.Oakley. and Stodola,R.King.
TITLE Compounds
JOURNAL Patent: US 6348328-A 219 19-FEB-2002;
FEATURES
source 1..791
location/Qualifiers
BASE COUNT 258 a 140 c 168 g 224 t 1 others
ORIGIN

Query Match 27.5%; Score 265.6; DB 6; Length 791;
Best Local Similarity 97.9%; Pred. No. 6.7e-39;
Matches 279; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 1 AGTAACCTTATCTCAAGAGTAGACATGAAGAAAAATTAATTTGGCAGTGCCATCACA 60
DB 507 AGTAACCTTATCTCAAGAGTAGACATGAAGAAAAATTAATTTGGCAGTGCCATCACA 566
QY 61 CTATTATCAGTAGCAACTTTAGCAGCTTGTGTGAAGGGTTCAGAAAGTGTCAACCTTATC 120
DB 567 CTATTATCAGTAGCAACTTTAGCAGCTTGTGTGAAGGGTTCAGAAAGTGTCAACCTTATC 626
QY 121 AGCATGAAGGGGATGTCTATACAGAAACATCAATTTATGACAGCAAGTGAAGAACCCCT 180
DB 627 AGCATGAAGGGGATGTCTATACAGAAACATCAATTTATGACAGCAAGTGAAGAACCCCT 686
QY 181 TCAGCCCAACAGTCTTGTAAATATGACATCCAAAAAGTTTGG-AAAAACAATATG 239
DB 687 TCAGCCCAACAGTCTTGTAAATATGACATCCAAAAAGTTTGG-AAAAACAATATG 746
QY 240 CTCAGAGCTTGATGATTAAGAGGTTGATGATATCTATGCGGAGAGA 284
DB 747 CTCAGAGCTTGATGATTAAGAGGTTGATGATATCTATGCGGAGAGA 791

```

```

RESULT 13
AE014908      12012 bp   DNA      linear   BCF 25-Oct-2002
LOCUS        Streptococcus mutans UA159 section 56 of 185 of the complete
DEFINITION   genome.
ACCESSION    AE014908 AE014133
VERSION      AE014908.1 GI:24377016
KEYWORDS
SOURCE       Streptococcus mutans UA159
ORGANISM     Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
              Streptococcus.
REFERENCE    1 (bases 1 to 12012)
AUTHORS      Ajdic,D., McShan,W.M., McLaughlin,R.E., Savic,G., Chang,J.,
              Carson,M.B., Primeaux,C., Tian,R., Kenton,S., Jia,H., Lin,S.,
              Qian,Y., Li,S., Zhu,H., Najjar,F., Lai,H., White,J., Roe,B.A. and
              Ferretti,J.J.
              Genome sequence of Streptococcus mutans UA159, a cariogenic dental
              pathogen
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 99 (22), 14434-14439 (2002)
MEDLINE      22295063
PUBMED      12397186
REFERENCE    2 (bases 1 to 12012)
AUTHORS      Ajdic,D., McShan,W.M., McLaughlin,R.E., Savic,G., Chang,J.,
              Carson,M.B., Primeaux,C., Tian,R., Kenton,S., Jia,H., Lin,S.,
              Qian,Y., Li,S., Zhu,H., Najjar,F., Lai,H., White,J., Roe,B.A. and
              Ferretti,J.J.
              Direct Submission
              Submitted (09-JUL-2002) Department of Microbiology and Immunology,
              University of Oklahoma Health Sciences Center, 940 St Young Blvd.,
              Oklahoma City, OK 73104, USA
FEATURES
    source
        location/Qualifiers
            1..12012
            /mol_type="genomic DNA"
            /strain="UA159"
            /db_xref="taxon:210007"
            95..1063
            /locus_tag="SMU.641"
            95..1063
            /locus_tag="SMU.641"
            /note="Best Blastp Hit: emb|CAB65629.1| (A1136149)
            putative oxidoreductase. [Streptomyces coelicolor A3(2)]"
            /codon_start=1
            /transl_table=11
            /product="putative oxidoreductase"
            /protein_id="AAN58375.1"
            /db_xref="GI:24377017"
            /translation="MKAVYIQTGSPATIVGSLPYPKYRPSVLIKVIAASINVDL
            FRSGIQDLPNPYILGRDAIGVAVSIAATVIVQIMELKPKOKLLEGAGHGSX
            ALIPERLEITPRKADSLKLGAVSAATVIVQIMELKPKOKLLEGAGHGSX
            LVLANEMGAEVVTVASLKSDFSRKLOKSPALCFIDKSLFKLSAYDGFPHIDT
            SEIPLQECDLAKRGVSLITAPKDSQDSROCYVMCOOKMGFVYIASHAKOLKKA
            GILLDAFROGGLLEEDVYIKFKDAAEHDMETKKEKRKILIVP"
            1308..1589
            /locus_tag="SMU.642"
            1308..1589
            /locus_tag="SMU.642"
            /codon_start=1
            /transl_table=11
            /product="hypothetical protein"
            /protein_id="AAN58376.1"
            /db_xref="GI:24377018"
            /translation="NNILDRYVSLDLAKRKRKLTLCNLFSENAITFANDGNTYSG
            REINPFEKFPYKNSLKHMTKEIADDMRQIGRIYAAKLESTLL"
            1695..2588
            /locus_tag="SMU.643"
            1695..2588
            /locus_tag="SMU.643"
            /note="Best Blastp Hit: prf|11718306A| xylanase
            [Butyrivibrio fibrisolvens]"
            /codon_start=1
            /transl_table=11
            /product="putative esterase"
            /protein_id="AAN58377.1"
            /db_xref="GI:24377019"
            /translation="MKINTSMLEWDKAKSEVYTYTYLIERKPDINKNRPRAVITCGG
            GGEFKVTDREKEPVALLFKNGFOATLTDYIVGKSGYRKPPLYLAKMLLYRENNON
            WNIDKEVYIFIGFSAGATASASLANSMNEPFLQKYPDYGETLKPGLVILSYPLDEN
            YOEQVSDPDNDMPKLSMPKPRIDFLTQALKTVYVGEKLTETKREKSPISHTSTATY
            PTFIMMODDCIYNVALLEYAKRLKDNILVLEHMFALGEHGLSYINSNTQAKPSDY
            FELGIKTKICIQFINRVHLK"
            2896..3834
            /locus_tag="SMU.644"
            2896..3834
            /locus_tag="SMU.644"
            /note="Best Blastp Hit: gb|AAC3746.1| (AF052209)
            competence protein [Streptococcus pneumoniae]"
            /codon_start=1
            /transl_table=11
            /product="putative competence protein/transcription
            factor"
            /protein_id="AAN58378.1"
            /db_xref="GI:24377020"
            /translation="MLVARDKNGKLLNLNGIPDKSDPFCPCACQSPVRLKNGWRPH
            FAVVALDCKEYSENESEAHNLNKAELVOSLSQTESVEYERKIPPEQIADLLVHNL
            ALEVEDCKSEARLCEKERTQAYQANGVQVLMLEKMLDRLRSLNHLKQFLYSQNGE
            HMEVDINRREIRLKYLIYEDIFGKYQYFTRSCSPGNTIMTFLQIAPAKOTLYSPVH
            ORRQVLAIOKOLARPNMLRQOELAYSQGRNLIIQSDADFPQVRLPQCSRGFCOI
            NODVSHFSAPFOYQOKKKEIQTLFPPAFYDKMK"
            3856..5655
            /gene="pepB"
            /locus_tag="SMU.645"
            3856..5655
            /locus_tag="SMU.645"
            /note="Best Blastp Hit: sp|O53778|PEPB_SRRAG GROUP B
            OUTGOPEPTIDASE PEPB >g|112786731|p|175178| chimeric
            oligopeptidase (EC 3.4.24.15) PEPB [validated] -
            Streptococcus agalactiae >g|11469939|gb|AAC44215.1|
            (U49821) group B oligopeptidase PEPB [Streptococcus
            agalactiae]"
            /codon_start=1
            /transl_table=11
            /product="putative oligopeptidase"
            /protein_id="AAN58379.1"
            /db_xref="GI:24377021"
            /translation="MSDNRSHTIEKYQMDLTSIPNDQWEKEKSLAKDLFAVRA
            GILLDSOSLLEITELXYLASRLEKLYVYASMKNDQDTVSYQEOIAKKAVALIAEA
            NOTPAPYEPPEMLSTEKYQDPLADCPDLLPIQHFPEKLFKKHVLSEKREALLAEA
            SEIFGAGETFEFLIDNADILIPVAVDETEEEVLSHGNFISIMSEKNRRVROEAVKAM
            VSIIEBOYQTVAKTLDQTNVKNVNFQAKIRNYVSARBAALANFIPETVYDVLVEAVN
            HPIPLQRYVYKQNVYLGDDLMKMYDLYVLSNDMVFYEALDLAAEVLALFEGRDYA
            EHVHRAFRERMDIRVYKNGKRSAGVSGSYDYNFAFLMWOQTLIDLPLFLVHETGSA
            HSTFEPRENDPYVYGDYSIFLAELASTNTNLIETLTKATDEKORFALLNHYLDGFK
            GTVFQTOFAFEEDDILYKADQAGVLTSPYLNLIADLLEKRYLGSKSNENPDIQYEMA
            RIPHRYNYFYVYQATGFAASFLAKHYHGNQEDKDKLTNLTGKANSQDPYLVAYAKA
            GUDWTESTYLDQAFVFEERLLELEVERKID"
            5868..6485
            /locus_tag="SMU.646"
            5868..6485
            /locus_tag="SMU.646"
            /note="Best Blastp Hit: p|11A57362|gyrB protein -
            Streptococcus pneumoniae (fragment)
            >g|11052803|emb|CA58770.1| (X83917) orf1gyrB
            [Streptococcus pneumoniae] >g|11490396|emb|CAA91552.1|
            (267740) unidentified [Streptococcus pneumoniae]"
            /codon_start=1
            /transl_table=11
            /product="putative phosphatase"
            /protein_id="AAN58380.1"
            /db_xref="GI:24377022"
            /translation="MPQAFIMWIDIGTLTDSYEAIMEALAVYTRTFGPFQAPRIARY
            IIKESVGKLDILAKOHGSLSEELKTFPIQEOYKKDKDKIKMLPPAKELQWAKDKGKIK

```

[illegible]

Oy	227	AAAAACAATATGCGTCAGACGGTTGATATGATAAAGAGGTGGTGAATACATATTGCCAACAA	286
Dd	7506	AAAACAATACGGCGGATTAAGCTTTCTATATAGAAGTAGCAAAAAGCTTTATATGAAGCTG	7565
Oy	287	AAAAACAATATGCGGGAAGAACCTACCACGCCTCTTGTCACAAGCAGGTATGACTCTTGAAA	346
Dd	7566	CTAGTACTATGATGATATCTTTCTCCAGTGGTTTTGGCTTACCGGGGCTACACAAAAGAG	7625
Oy	347	CACGTAAGCTCAAATTGTCACAACTAATTAGTTGAGTTGGCAGTTAAGAGSTACAG	406
Dd	7626	ATTATTAAGAAAACAATTGCTTCTGAAAAATTAAATAGATATATGCTTTAAGGAAGAACTA	7685
Oy	407	AAGCTGAATTACAGATGAAGACCTATTAAGAAGACCTTGATGATACACTCCAGATGTAA	466
Dd	7686	AGAAAAGAAATACAGACGCTAGTTATAGTCAGCTTATTAAGGATTATTAACCGGAAGTAA	7745
Oy	467	CGGCTCAATCATCCGCTTAAATATATAGAATAGGCCAAAGAAGATTCTGCAAAAAGCCA	526
Dd	7746	CAGCTCAAGTATTCATTTGGATAGTGAAGATTAAGCAAAAATCTGTTTAGAAGAACTA	7805
Oy	527	AGGCAAGAGTGTGATTTTGGCTCAATTAAGCCAAAGATTAATCACTGATGAAGAAAACAA	586
Dd	7806	AGGCAGATGTGCGACAGCTTGCTAAATTTGCTAAAGATTAAT-----ACCA	7850
Oy	587	AAGAAATNGTGGAGAAATTAATCTTGATTTGCTTACACAGAAAGTACCTGAGACATCA	646
Dd	7851	AAGAGATAGACCGAATATATAGCTTTGATAGGCTCAACTAACTTCTGATGAAGCTTT	7910
Oy	647	AAAAAGCGCGCTTTCGCTTATAGATGATGGATGGTCTTTCGATGATATTAACAGCACTGCA	706
Dd	7911	TATGGCGTCTTGAATCTTGATTAAGAGTAGGCGTTTCAGATGTCATTAAAGCGCTGAT	7970
Oy	707	CACAAGCTACAGTAGCCATATTACATTTGTAATGTAACCTCACTAAGAAAAACAGAAATCAT	766
Dd	7971	CAAGACTTATAAACCTGTTACTATATTTGTCAAAAATCACGAAAAAGCATGACAAAAAGC	8030
Oy	767	CTAATATTGATGACGTACAAAAGAAAATTAATAAACCTGTAATTTCTGACCAAAAACAAAT	826
Dd	8031	CCGATTTGGAAGCTTATTAAGAAAGCTGTGAAAGAAATTAATGCTTTCCAAAATTTAAATG	8090
Oy	827	ATTCACACTTTGTTCAAAAGCATTTACGAAAAAGATTTGCAAGCAGCCATATTCAGGTTA	886
Dd	8091	ACAGTAATTTCCGTAATGCTGTATTATGGAAGAACCTTTAAGAAAGCTAATGTCAAAATTA	8150
Oy	887	AGACCAAGCTTCCAAAATATCTTTTACCCCAATAT	921
Dd	8151	AAGATTAAGCGCTTTAGTGAGATCTTATCACAAATAT	8185

RESULT 14
AE006576/c LOCUS AE006576
DEFINITION Streptococcus pyogenes M1 GAS strain sp370, section 105 of the complete genome.
ACCESSION AE006576
VERSION AE006576.1
KEYWORDS GI:13622488
SOURCE Streptococcus pyogenes M1 GAS
ORGANISM Streptococcus pyogenes M1 GAS
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.

REFERENCE
1 (bases 1 to 11615)
Ferretti,J.J., McShan W.M., Adjić,D., Savic,D., Savic,G., Lyon,K., Primeaux,C., Sezate,S.S., Suvorov,A.N., Kenton,S., Lai,H., Lin,S., Qian,Y., Jia,H.G., Najjar,F.Z., Ren,Q., Zhu,H., Song,L., White,J., Yuan,X., Clifton,S.W., Roe,B.A. and McClaughlin,R.E.
Complete genome sequence of an M1 strain of Streptococcus pyogenes Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)

TITLE JOURNAL MEDLINE PUBMED
11296296
2 (bases 1 to 11615)
Ferretti,J.J., McShan,W.M.,Adjić,D., Savic,D., Savic,G., Lyon,K.,

TITLE

Primeaux, C., Sezate, S.S., Surorov, A.N., Kenton, S., Lai, H., Lin, S., Qian, Y., Jia, H.G., Najjar, F.Z., Ren, Q., Zhu, H., Song, L., White, J., Yuan, X., Clifton, S.W., Roe, B.A. and McLaughlin, R.E.
 Direct Submission
 Submitted (10-Apr-2001) Department of Microbiology and Immunology,
 University of Oklahoma Health Sciences Center, 940 St Young Blvd,
 Oklahoma City, OK 73104, USA

FEATURES

source

1. 11615

/organism="Streptococcus pyogenes M1 GAS"

/mol_type="genomic DNA"

/strain="SE370"

/serotype="M1"

/db_xref="taxon:160490"

/complement(77. .772)

/gene="Spy1384"

/complement(77. .772)

/gene="Spy1384"

/note="Best Blastp hit = p1r11576167 hypothetical protein

- Synecocystis sp. (strain PCC 6803)

>g11653513|dbj|BA18426.1| (D90914) hypothetical protein

[Synecocystis sp.]

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAK34205.1"

/db_xref="GI:13622489"

/translation="MKGFINYLKIAVLIITLAWENVLPWILKOKHIDIPVLMWGIGI

FLIVIVGSVLIYMLGYOKKODPTIKOOKMRLVDMGYLAPLILTRVIAVGTINOL

WGGQOVSANDAIHRLIARIKGEPFLYTLFVYLFARIPMEELVFRFPMDLEFG

KSLKVAGLTSLVFALPHATNSVEFLTMSCKGIFLVAYORGNLADALLHIFNLI

EYILMLSIGLGI"

/complement(791. .1552)

/gene="Spy1385"

/complement(791. .1552)

/gene="Spy1385"

/note="Best Blastp hit = no blast hits"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAK34206.1"

/db_xref="GI:13622490"

/translation="MKKKNSPLRLKMLSSLAGGITGAWGAFGLYHGRDLHIT

FLKDDVINILILNLNVVDTLPLSVFLTLQKKEFAVNTIEEDDISNGYKOLKK

HAYTMLIIVASTLSMCNVLLGTLTINDSOHMLAIPLDILLMLVTFPOALAMRY

NAIRGTDPVYFNNKELKHNIALDEAELOAHKHSFESVLSLNGVILPSLYVILEFV

YLFTGOVELTALIVLVLIDLLVKSATMTROFYR"

/complement(1549. .1764)

/gene="Spy1386"

/complement(1549. .1764)

/gene="Spy1386"

/note="Best Blastp hit = sp1028481|YH93_ARCFU HYPOTHETICAL

TRANSCRIPTIONAL REGULATOR AF1793 >g117483936|pir||H69473

repressor protein homolog - Archaeoglobus fulgidus

>g112648767|gb|AA89470.1| (AF000979) repressor protein

[Archaeoglobus fulgidus]"

/codon_start=1

/transl_table=11

/product="putative transcriptional regulator protein"

/protein_id="AAK34207.1"

/db_xref="GI:13622491"

/translation="MEVILKNRLKELRANDGINOTEMAKLAGVROTISLIERNEYTP

SVIAMKIAKVEPEPEVEEVRIVEE"

/complement(2125. .4743)

/gene="alals"

/note="synonym: Spy1389"

/complement(2125. .4743)

/gene="alals"

/EC_number="6.1.1.7"

/note="Best Blastp hit = sp1034526|SYA_BACSU ALANYL-TRNA

SYNTHETASE (ALANINE-TRNA LIGASE) (ALARS)

>g117437617|pir||A69584 alanine-TRNA ligase (EC 6.1.1.7)

alals - Bacillus subtilis >g112635186|emb|CAB14682.1|

(Z99117) alanyl-trna synthetase [Bacillus subtilis]

>g112635204|emb|CAB14699.1| (Z99118) alanyl-trna

synthetase [Bacillus subtilis]"

/codon_start=1

/transl_table=11

/product="putative alanyl-trna synthetase"

/protein_id="AAK34208.1"

/db_xref="GI:13622492"

/translation="MKELSAQIRQMIDPFKSGHCVEPSANLVPVNDPTLWINS

VATLKYPDGSVTPENPRITNAOKSIRPNDIENVGKTRHHTMEFLNFSIGDFRD

EALFEGFELTSPDDEFDPKDTWTVYDDKTSNRIJAGNIVSLHSGNDPAPRS

GAGSGPPTETFPDGEDPDENITRLIADIDENDRIETLNTVLSQFNADPAPRS

EKELPNKNIIDPAGLERLAAMQAKTNFETDLMPIIREYKLSGTTYPPDDNMS

EKVIDHIALSFAIGDALPGENEGYVLRRLRAVHGRRLGINTETLKLPVY

GOIMESYVPEVLEKRDPTKIVREBEFTARTIDAGSHDSLAAQLAEKDLLEGK

DIFKLYDYGFVEVLEELAEADQKIDHEGKFSAMKEQODRARAAVKGSGMGN

TLAGIIEESREFDYSLSSLSVITIAINEREAVSEGOALLVFPQTFVAMEGQVA

DTGRIRKNGDPAFVAVYOKAPNGOPIHTVAVLASLVGNVYTLIEIKERRIAVEK

HTATHLAALHNVIGEHAIOGSLNEEFLRFDTHREAVSNEELRIEDEVNQIV

NALTTTETEDVETAKEGAMALFEKYGKAVRVQIGNYVSELGTHLNSSIEIGL

FRIVEEIGSGSTRRIIVTGRQAEAVRNQEDALKEIAATVAKAPQLDAAKVQALS

DSLRLQKENAELEKKAAMAAAGDVFEKQKGVGFYASGVDAADALRTFFDNMK

OKDYSADVIVVAIGEKVNLVAASKTQDVHGNMKEIAPVYAGRGSGKPPMAAGGS

DASKIAELLAAYAEIV"

/complement(5130. .6185)

/gene="Spy1390"

/complement(5130. .6185)

/gene="Spy1390"

/note="Best Blastp hit = sp115294|PRTN_LACLA PROTEASE

MATURATION PROTEIN PRECURSOR >g11799781|pir||S08083

probable protein export protein prtm precursor --

Lactococcus lactis subsp. cremoris (strain NCDO 763)

plasmid pLp763 >g1147198|emb|CA32349.1| (X14130) ORF (AA

1 to 299) [Lactococcus lactis subsp. cremoris]"

/codon_start=1

/transl_table=11

/product="putative protease maturation protein"

/protein_id="AAK34209.1"

/db_xref="GI:13622493"

/translation="MKNSMKLIVSVYTLASVALAACOSTNDNTKIVSMKGGTISVD

FYENKTEEVSOAKMLNIVISRFPAQVGDKSKREVEKATKEQAGASALAAQ

SSLPTPEETROIRSSKLIVEYAVEKAEKELTTOEYKKAESYTPMAVEMITLDEET

AKSVLEELKAEAGADFTALAKEKTTPEKKVTVYKPPSGATNPTDVVAKASSLNEGIS

DVISVLDPTSYOKKFIYVVTYKAKKSKSMQYKKRIRAKIITIAESKMNPNQKYIAN

AEDPTEAQTEPPAAQ"

/complement(6248. .6955)

/gene="Spy1391"

/complement(6248. .6955)

/gene="Spy1391"

/note="Best Blastp hit = gb1AAC46080.1| (AF016233)

methyltransferase [Enterococcus faecalis]"

/codon_start=1

/transl_table=11

/product="putative methyltransferase"

/protein_id="AAK34210.1"

/db_xref="GI:13622494"

/translation="WVKSYSKTANHNMRPVPVKEELVHYMTRORQOTGFLAELOFA

ROBNPIIOPEVAVYRFLQSLQPKHIEITAGLSLMAENAPDATTIYIDRNR

EMIDPAKNAFAKVSROQIRLEGGADAILSTLEGDFDVFDMSAKSVIYLPETLR

ILKVGQVTLIDVFOGSDITKPIEDIRROQRITYRQSLPFPATLTHRLTISLVPLS

DGLMKRKQADIVLPD"

/complement(7021. .8217)

/gene="Spy1392"

/complement(7021. .8217)

/gene="Spy1392"

/note="Best Blastp hit = dbj|BA806247.1| (AP001515)

unknown conserved protein in others [Bacillus halodurans]"

/codon_start=1

/transl_table=11

/product="putative oxalate:formate antiporter"

/protein_id="AAK34211.1"

/db_xref="GI:13622495"

```

/translation="MEKTKRYIIATAGILLHLMIGSTYAMSVENPILQETGMDQAPV
AFAPSLAFCILGSLAFMGNLVEQYPRLTSTVSAIILASGNMLTGLAIDKEILMLY
IGVAVIGIGAGYITPISITIKMPDKRMATGAPLMSGFPISLISPTAOLMLET
EGVATFYILGLIYLIMLPASOLIIKPTAEIAILDKRKIÖNNNSYILEGRTAKALK
TKSTYCLMWILFIIITGGLISVAPAPADLTGMSPEMSIIVGAMGTFEGRLVW
ASLSDYIGRATVILFLVSLIMTISLIFAMSLIFMISITMTLTCYAGAGSLIPYL
SDLEGAKEELTLHGVIILTAWAIIAALTGMLISITVEMTHNLLTLCVEFIVLYIILMW
ALRLKK"
gene
complement(8593..10398)
/feature="pep"
/note="synonym: SPY1393"
complement(8593..10398)
/feature="pep"
/EC-number="3.4.24.15"
/note="Best Blastp hit = sp|P54124|PERL_LACIA
OLIGOENDOPEPTIDASE F, PLASMID >gi|2126616|pir|I1A55485
11goendopeptidase F - Lactococcus lactis
>gi|510140|emb|CAA8354.1| (232522) 11goendopeptidase F
[Lactococcus lactis]"
/codon_start=1
/transl_table=11
/product="putative oligopeptidase"
/protein_id="AAK34212.1"
/db_xref="GI:13622496"
/translation="MTDNRSHLIEEKYTWDLSTIFATDKMEAEVSDIATEVEASKGEA
GHLDSSANLTKVTKYLELARVEKYVYVYHMKNDOPITVAKYOEYQAKAGSLYAKF
SEVSPYDEYVMLEHEDYQAFLETPELKYNHPEKLFQAREHVLISQARELLAGA
EITNGAEFEPSITLNDIYPPVYKNDKGEVETLHGFIISMSKDSQVROAYEAM
KSTYEOFOHTYAKTLQTNKYNQKARHKIKDSRQAMAMNFIPEAYITDLELVNK
HLPLHRLKLRQVELGLDLMKMDVYTPLETDLAIGDYDALKEAVLIVAFKDYA
DLVRAETFERMIDIVHNKGRSGASGGSYDTNFIILNMDDITLNTLILHGHSL
HSFTRETOPYVGDYSIFLAEIASTMENTNALNEVODEKEREFITLHGHSL
GTVRQTOFAFEHAIHQADQGEVLSEVINOIYADINEKYVIGLSKDNFIQYEMA
Query Match 27.1% Score 261.6; DB 1; Length 11615;
Best Local Similarity 56.9%; Pred. No. 2e-38;
Matches 505; Conservative 0; Mismatches 374; Indels 9; Gaps 1;
34 AAAAAATATATGAGAGTGCATATATATGAGAGAGCACTTTGACAGCTTGTCG 93
Db 6173 AATTAACATCATTCCTAGTGTGTCATTCGCTCAGTACGATGCGCTTTGACAGCTTGTCAA 6114
QY 94 AAAGGCTCAGAGGTCAGACCTTATCAGATGAAGGAGATGTCATTAACAGAACATCAA 153
Db 6113 TCACTAATGACATTAAGTATATTCGATGAAGAGTGATTAACATTAAGGTTATGATAT 6054
QY 154 TTTTATGAGCAAGTGAAGCAACCCCTTCAGCCCAACAAGTCTTGTAAATATGACCATC 213
Db 6053 TTTTACATTAAGCAACCAACCAACAGAAATATCGCAAAAGGAGTCTTAATCTGGTAAT 5994
QY 214 CAAAAAGTTTGAAGAAACATATGCTCAGAGCTTGATGAATGAAGAGTGTGATGATCT 273
Db 5993 AGTGTGTTTGAAGCTCAATATGCTATAGGTTTCAAAAAAGAAAGTGTGAAGAGCG 5934
QY 274 ATGCGCGAAGAAAAAACAATATGCGCAAAACATCAACAGTGTCTTTCACACAGCAGT 333
Db 5933 TATCATTAACAGCTGACAGATGCGCTTCATCTCTGCGCTTGGACCAATCAAGC 5874
QY 334 ATGACTCTGGAACACGTAAGCTCAATTCGTACAAATTAATGATGAGTGGCAGTT 393
Db 5873 TTGACACCTGAGACTTTTAACGCTCAGATCGCTCTTCAAAAATGATGAATATGCGGTT 5814
QY 394 AAGAAAGTAGAGAGAGTGAATGACATGAGAGCTTAAGAAAGCCTTGTGATGATGATC 453
Db 5813 AAAGAGAGAGCTAAAAAAGATATGACACACAGAAATATTAAGAAAGCATATGAATCTTAT 5754
QY 454 ACTCCAGATGTAACGCTCAATCATCCGCTTAATTAATGAAGATAGAGCCCAAGAGATT 513
Db 5753 ACTCCAAATATGCGAGTGAATGATTAATGATTAAGAGAGAGAGAGAGCTTAATCAAGTC 5694
QY 514 CTCGAAAAAGCCAGGAGAGAGTGTGATTTTGTCTCAATTAAGCCAAAGATTAATCAACT 573
Db 5693 TTAGAGAACTTAAGAGCGAGAGAGCTTTACAGCTATTTGCTTAAGAAAAAACAAAC - 5635

```

```

QY 574 GATGAAAAACAAAAAAGATGTCAGAAATTAACCTTTGATTCGTTCACAGAACTA 633
Db 5634 -----AACACCTGACAAAAAAGATGACCTATTAATTTGATTCAGGTCACAAATTA 5583
QY 634 CCGTAGCAAGTCAAAAAAGCCGCTTCGCTTTAGATGATGATGATGATGATGAT 693
Db 5582 CCGAGTATGTCGTAAAAAGCGGCTTCAGATTTGATGATGATGATGATGATGATGAT 5523
QY 694 ACAGCACTGGCAGCAACCAAGCTTACAGTACCAATTTTCAATTTTCAATTTTCAATTT 753
Db 5522 TCGGTTTATGATTCACACTTCTTATCCAAAGAGTTTACATGTTTAAAGTCAATTA 5463
QY 754 ACAGAAAAATCATCATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 813
Db 5462 GCAGAAAAAATTCAGATTTGCGCAGAAATTAATGAAGAGCTTTGAAGAGCTATTAAT 5403
QY 814 CAAAAACAAAAATGATTAACATTTGTTTCAAGCAATTAACGAAAGAAATTCAGAGCACC 873
Db 5402 GAAAAATCAAAAAGATGATTAATTTCCAAACAAAGTTTATTCGAAATGATGATTAAGCT 5343
QY 874 AATATCAAGTTAAGAGACCAAGCTTCACAAATATCTTACCCATAT 921
Db 5342 AATGTAATAATTAAGACAAAGCTTGTCTAATATTTGGCGCAATAT 5295

RESULT 15
AE014157 51454 bp DNA linear BCT 19-JUL-2002
LOCUS Streptococcus pyogenes MGAS315, section 22 of 37 of the complete genome.
DEFINITION AE014157 AE014074
ACCESSION AE014157.1 GI:21904768
VERSION
KEYWORDS
KEYWORDS
ORGANISM
REFERENCE
1 (bases 1 to 51454)
Beres,S.B., Sylva,G.L., Barbican,K.D., Lei,B., Hoff,J.S.,
Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F.,
Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and
Musser,J.M.
Genome sequence of a serotype M3 strain of group A Streptococcus:
phage-encoded toxins, the high-virulence phenotype, and clone
emergence
Proc. Natl. Acad. Sci. U.S.A. 99 (15), 10078-10083 (2002)
2 (bases 1 to 51454)
Beres,S.B., Sylva,G.L., Barbican,K.D., Lei,B., Hoff,J.S.,
Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F.,
Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and
Musser,J.M.
Direct Submission
Submitted (14-JUN-2002) Laboratory of Human-Bacterial Pathogenesis,
Rocky Mountain Laboratories, NIAID, NIH, 903 South Fourth St.,
Hamilton, MT 59840, USA
LOCATION/Qualifiers
1. 51454
/organism="Streptococcus pyogenes MGAS315"
/mol_type="genomic DNA"
/strain="MGAS315"
/serotype="M3"
/db_xref="taxon:198466"
/note="group: A"
complement(150..1139)
/feature="grab"
/feature="grab"
/note="synonym: SPYM3_1032"
complement(150..1139)
/feature="grab"
/note="best blastp hit: gb|AA026342.1|AF124403.1
(AF124403) GRAB precursor [Streptococcus pyogenes]"
/codon_start=1
/transl_table=11

```

```

/product="protein G-related alpha 2m-bindingprotein"
/protein_id="AAM79639.1"
/db_xref="GI:21904769"
/translation="MGKEIKVYFLRBSAFLVAVASVIVSGTVSAVSPPEIPRII
PNGRTLTNLGNAPKALRNEREAIIDELKQATIDKREATTIAEASSDALEALDDQA
DALOSEEAAYVOSTAASDADLEALADQTDALQSEEAAYVADTAASDALEALADDTQ
VQSEASVYKADTAASDALEALADQTDALQSEEAAYVADTAASDALEALADDTQALQ
SEEAAYVADTAASDALEALADQTDALQSEEAAYVADTAASDALEALADDTQALQ
TKDTPVYVKKERQVNTLPTTGSESNPEFTAAALAIWSTGALVSSCKEN"
complement(1496..2710)
/gene="murZ"
/note="synonym: Spym3_1033"
complement(1496..2710)
/gene="murZ"
/note="best blastp hit: gb|AAL97966.1| (AE010057) putative
UDP-N-acetylglucosamine 1-carboxyvinyltransferase
[Streptococcus pyogenes MGAS8232]"
/codon_start=1
/transl_table=11
/product="putative UDP-N-acetylglucosamine
1-carboxyvinyltransferase"
/protein_id="AAM79640.1"
/db_xref="GI:21904770"
/translation="MAVSGAKNSVALIPAILADYILDCVPAISVDLSIETMEL
MGATVNHGDTLEIDPRGVODIPMPYKINSIRASYIYYGSLGREGQAVVGLPGCD
LGPRIDILHLAEAMGVSEYEGENNNLSTNGKTIHGAHITMDVSGATINTVAA
TKAOGTIVEMNAREPEIIVATILNMGAIHREGTIDITIOGVOKLHTRHYIPD
RIEAGTYIALAAIGKGVKINVTLEHSEFIAKLEEMGVMTVEDALIEVKEOSKL
ALTIKTSYPGPATDLOQPLNPLIKADRGRTIIDTYEKRNHYPELMRMGADISVI
GGQIYVQSPRLTGQVAKATDRAGAAVYTAGLMAEGKTEITNIEFIIRGASTIAKL
TALGADLIQIED"
complement(2929..4125)
/gene="mekT"
/note="synonym: Spym3_1034"
complement(2929..4125)
/gene="mekT"
/note="best blastp hit: gb|AAL97967.1| (AE010057)
S-adenosylmethionine synthetase [Streptococcus pyogenes
MGAS8232]"
/codon_start=1
/transl_table=11
/product="putative S-adenosylmethionine synthetase"
/protein_id="AAM79641.1"
/db_xref="GI:21904771"
/translation="MSERKLFTSESVSEGHDPKIDQISDAITLAILDEBEAHVAE
TCVYTGSVHVEGEISTYAIIDINRVVRDTIAIGTEAEYGSASVGVSHLSVQSG
DIAOGVNALRESREGDIDLSHIGADQGLMEGAINETPELMPLPISLSHOLYERLA
ELRSGEISYLRPDAKSQVTEYDEHDKPVRVDYVISTOHDPREATNDQIOTDYEKY
IKAVIPADYDDDKFEFIPNTPGRVPIYGGPODGLTGKTIIVDTYGGYSRRGGAFSG
KDATKVDASASTARIKATKNLYAGLAKAEYOLATAGVADPVSVRVDFTGISTVPE
AVLEAAVQVDFDLRPAGIITQMLDKRPYIKOTPAAYGHMGRTDIDLPWEHLNKVDALVE
AVKTVL"
complement(4662..7040)
/gene="inla"
/note="synonym: Spym3_1035"
complement(4662..7040)
/gene="inla"
/note="best blastp hit: gb|AAK34188.1| (AE006574) putative
internalin A precursor [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative internalin A precursor"
/protein_id="AAM79642.1"
/db_xref="GI:21904772"
/translation="MKTAKVILVGLLSQTLTACOSRGNTYPIKTKOSRKGMTS
NKIRPKSKTKNTKNGVAGVDEPTDGGFILTQSKILSKTDQIIVDHDGSHFIF
YADLKSPFEYLIPKASLAPVAAQAAQASQTSKVDAPHHHEPNPADIVAEADLQY
TVRHDHRYILKSLSSQTOQAQAVATRIPOSSPYSTATANGICPLHPTPSGRO
FNGGIVGVTQDNILVHDGHDPISTFIDKQGAHVAADQIDEPKAKAEPAETHHTP
ELSEREKYQERLAVIAEKIDIPSTIKRVEYODKGLGEPHHDHAAVLMISDEIG
KDIIDPHAIIEHARELEKHFQMDTLRALGDEEVLIDIVRTKIDPAPPSEKEDNM
KEMLATIVIKLIDGSRKDPLOKRGSLPLNLEFVIGGIPRIDISPVLOFKKIKOLML
KTGYTDVRFIDNMPLQEGIDISQNNLKDIFSLSKKNLTIVAAADNGIEDIRSLQLP

```

```

/gene="bira"
/note="synonym: Spym3_1036"
7244..8185
/gene="bira"
/note="best blastp hit: gb|AAK34189.1| (AE006574) putative
biotin operon repressor [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative biotin operon repressor"
/protein_id="AAM79643.1"
/db_xref="GI:21904773"
/translation="MKTSEKIYQLLSQTDDEYVSGEYLAQDLSISFTSVKSIKLENO
GIQIDSLKHKGYRMVQGDILPRTISQIGMPVITYPHQSOTOLDPAKQIEAHNSAPR
LYLAPSOBAKGRDLROFFSASTGCIYMSMYLKPVPVADMPPTMVAASSIVKAIIR
LTGIDTEIKWVDIYLGNNHKAAGILTEAITSVEGLTIDYIIGVGLNFVDEPEAIA
OKAGSLFTEKEPITRNLDLIDIKRFLSIPVADHYKVKYKXSIVLNKQVTFIENSQEK
RAIADITDQGLIYQFENGDIQTRSEISLSMEN"
complement(8160..8444)
/gene="Spym3_1037"
complement(8160..8444)
/gene="Spym3_1037"
/note="best non-GAS blastp hit: gb|AAK74993.1| (AE007392)
hypothetical protein [Streptococcus pneumoniae TIGR4], and
gb|AAK99574.1| (AE008452) hypothetical protein
[Streptococcus pneumoniae R6]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAM79644.1"
/db_xref="GI:21904774"
/translation="MGKVLPLETFYLVTFEHSYFECYKAWTIRQFLMAFVCAFEF
YFFNDLLSGNVLFLFWGLLFRDLRVRHTINQTLTKILTANSPKKRD"
complement(8448..10118)
/gene="dnaX"
/note="synonym: Spym3_1038"
complement(8448..10118)
/gene="dnaX"
/note="best blastp hit: gb|AAL97971.1| (AE010058) DNA
polymerase III gamma/tau subunits [Streptococcus pyogenes
MGAS8232]"
/codon_start=1
/transl_table=11
/product="DNA polymerase III gamma/tau subunits"
/protein_id="AAM79645.1"
/db_xref="GI:21904775"
/translation="MYQALYRKYSQTFDEWVGQSVISTTLKQAVESGKISHAYLFSG
PRGTSKTAAKIFAKAMCNPNOVDEPCNOCIDITNGSLSEYIEIDAASNNQVDE
IRDIRDKSTVAPSRATYVYIIDEVHMLTKEGAFNMLTLEPPTENNVYIATTELHK
IPATILSRVOREPEAKIKOKAIRELAVYLDKEGAYVADANLITARRAREGMRALS
IILQALSTSPDNQVTTAIEETIGTISMLADIVRYVSOEDATDALALETTIDSGK
MSRFRATDLITYLRDLVAKAGDNQASAVFDNLSTSIDIREQMITVYVSHLSEIK
KGTBRPIAEMWTIOIAQEQIOLSQANISGELISIEITELKELAQLOKQOLSQSRPD
SLARSDKTKPTASVRYRVITLIKMEETVRSQOSROYLDLAKNANMEIIDNISADQ
RALMGSEPVLANSENATILAEFAANAOVMSRNMLNMFQIMSKAGFSPNLAIVP
RTDPQHIKKEFAQCKSOKDSVQEQEVALDIPEDFLDLKINTID"
complement(10118..10681)
/gene="Spym3_1039"
complement(10118..10681)
/gene="Spym3_1039"
/note="best non-GAS blastp hit: gb|AAK06294.1| (AE006448)
hypothetical protein [Lactococcus lactis subsp.
lactis]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAM79646.1"
/db_xref="GI:21904776"

```


BEST AVAILABLE COPY

THIS PAGE BLANK (USPTO)

PT pneumoniae useful for treating or preventing microbial infections,
 PT specifically streptococcal infections
 PS Disclosure: Fig 1: 18pp; English.
 XX Streptococcus pneumoniae is a major cause of invasive diseases such as
 CC meningitis, bacteraemia and pneumonia, as well as non-invasive diseases
 CC such as acute otitis media and sinusitis. The present invention relates
 CC to a vaccine comprising Protease Maturation Protein (Pmp) of
 CC S. pneumoniae. The vaccine can be used for the treatment of S. pneumoniae
 CC infections. The present sequence is the coding sequence of S. pneumoniae
 CC Pmp, which can be used to produce the vaccine of the present invention.
 CC Pmp is a surface-associated protein.
 CC
 XX Sequence 966 BP; 374 A; 170 C; 187 G; 235 T; 0 other;
 XX
 SQ
 Query Match 100.0%; Score 966; DB 22; Length 966;
 Best Local Similarity 100.0%; Pred. No. 4.2e-213;
 Matches 966; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTAACTTATCTCAAGAGTAGACATGAAGAAAAATTTATGAGAGTCCATCACA 60
 DB 1 AGTAACTTATCTCAAGAGTAGACATGAAGAAAAATTTATGAGAGTCCATCACA 60
 QY 61 CTATTATAGTAGCAACTTTAGCAGCTTGTGGAAGGGTCAAGAGTGCAGACCTTATC 120
 DB 61 CTATTATAGTAGCAACTTTAGCAGCTTGTGGAAGGGTCAAGAGTGCAGACCTTATC 120
 QY 121 AGCATGAAGGGGAGTGCATTTACAGAACATCAATTTTATGACAGTGAAGAACCAACCT 180
 DB 121 AGCATGAAGGGGAGTGCATTTACAGAACATCAATTTTATGACAGTGAAGAACCAACCT 180
 QY 181 TCAGCCCAACAAGTCTTGTAAATATGACCATCCAAAAAGTTTGAAGAACATATGCG 240
 DB 181 TCAGCCCAACAAGTCTTGTAAATATGACCATCCAAAAAGTTTGAAGAACATATGCG 240
 QY 241 TCAGAGCTTGATGAAGAGGTGATGATGATGATGATGATGATGATGATGATGATGATG 300
 DB 241 TCAGAGCTTGATGAAGAGGTGATGATGATGATGATGATGATGATGATGATGATGATG 300
 QY 301 GAAACACTACCAAGCTGTCTGTACAGACAGGTATGACTCTTGAACACGTAAAGCTCA 360
 DB 301 GAAACACTACCAAGCTGTCTGTACAGACAGGTATGACTCTTGAACACGTAAAGCTCA 360
 QY 361 ATTGCTACAGTAATTTAGTGTGAGTGGCAGTTAAGAGGTAGCAGAACTGAATTTGACA 420
 DB 361 ATTGCTACAGTAATTTAGTGTGAGTGGCAGTTAAGAGGTAGCAGAACTGAATTTGACA 420
 QY 421 GATGAAGCCTATTAAGAAAGCCTTGTATGATGATGATGATGATGATGATGATGATGATG 480
 DB 421 GATGAAGCCTATTAAGAAAGCCTTGTATGATGATGATGATGATGATGATGATGATGATG 480
 QY 481 GCTCTTAATTAATGAAGATGAAGCCCAAGAGTTCGAAAAAGCCAGGAGAGGTGCT 540
 DB 481 GCTCTTAATTAATGAAGATGAAGCCCAAGAGTTCGAAAAAGCCAGGAGAGGTGCT 540
 QY 541 GATTTGCTCATATGACCAAGATATATTCACCTGATGAAAAACAAAGAAATGCTGA 600
 DB 541 GATTTGCTCATATGACCAAGATATATTCACCTGATGAAAAACAAAGAAATGCTGA 600
 QY 601 GAAATTTACCTTTGATTCGCTTCAACAGAGTACCTGAGCAAGTCAAAAAAGCCCTTTC 660
 DB 601 GAAATTTACCTTTGATTCGCTTCAACAGAGTACCTGAGCAAGTCAAAAAAGCCCTTTC 660
 QY 661 GCTTTAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720
 DB 661 GCTTTAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720
 QY 721 AGCCAAATTTACATGTAATACTCACTAAGAAAAAGAAAAATCATATATTTATGAC 780
 DB 721 AGCCAAATTTACATGTAATACTCACTAAGAAAAAGAAAAATCATATATTTATGAC 780
 QY 781 TACAAGAGAAAAATTAATAAAGTGTATCTTGACTCAAAAAACAAATGATTCACATTTGTT 840

DB 781 TACAAGAGAAAAATTAATAAAGTGTATCTTGACTCAAAAAACAAATGATTCACATTTGTT 840
 QY 841 CAAGCATTATCGGAAAGAAATTCAGACGACCAATATCAAGTTAAGACCAAGCCTTC 900
 DB 841 CAAGCATTATCGGAAAGAAATTCAGACGACCAATATCAAGTTAAGACCAAGCCTTC 900
 QY 901 CAAATATCTTTACCAATATATCGGTGGTGGAGATTCAAGGTCAGAGTGTATCATCA 960
 DB 901 CAAATATCTTTACCAATATATCGGTGGTGGAGATTCAAGGTCAGAGTGTATCATCA 960
 QY 961 AACGAA 966
 DB 961 AACGAA 966

RESULT 2
 AAV52241
 ID AAV52241 standard; DNA; 11309 BP.
 XX
 AC AAV52241;
 XX
 DT 23-OCT-1998 (first entry)
 XX
 DE Streptococcus pneumoniae genome fragment SEQ ID NO:108.
 XX
 KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
 KW computer readable medium; vaccine; pharmaceutical composition; ds.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO9818931-A2.
 XX
 PD 07-MAY-1998.
 XX
 PF 30-OCT-1997; 97WO-US19588.
 XX
 PR 31-OCT-1996; 96US-0029960.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
 PI Kunsch CA, Rosen CA;
 XX
 DR WPI; 1998-272225/24.
 XX
 PT Computer-readable medium with recorded Streptococcus pneumoniae
 PT polynucleotide sequences - useful in diagnostic kits and assays, and
 PT pharmaceutical compositions and vaccines for Streptococcus
 PS pneumoniae
 PS
 PS Claim 1: Page 794-800; 1409pp; English.
 XX
 CC The present invention describes a computer readable medium which has
 CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
 CC recorded on it, or a representative fragment of a sequence at least 95%
 CC identical to SEQ ID NO:1 to 391. The nucleotide sequences depicted in
 CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
 CC Streptococcus pneumoniae. The present invention also describes an
 CC isolated nucleic acid molecule encoding a homologue of any of the
 CC fragments of the S. pneumoniae genome (SEQ ID NO:1 to 391) where the
 CC nucleic acid molecule is produced by a process comprising: (a) screening
 CC a genomic DNA library using as a probe a target sequence defined by any
 CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
 CC library which contain sequences that hybridise to the target sequence and
 CC isolating the nucleic acid molecules from the members; or (b) isolating
 CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
 CC molecules whose nucleotide sequence is homologous to amplification
 CC primers derived from the fragment of the S. pneumoniae genome to prime
 CC the amplification and isolating the amplified sequences. The computer
 CC readable medium can be used in a computer-based system for identifying
 CC fragments of the S. pneumoniae genome of commercial importance, or
 CC expression modulating fragments of the S. pneumoniae genome. Products

CC from the present invention can be used in diagnosis kits and assays, and
 CC pharmaceutical compositions and vaccines for *S. pneumoniae*.

SQ Sequence 11309 BP; 3557 A; 2073 C; 2451 G; 3228 T; 0 other;

Query Match 100.0%; Score 966; DB 19; Length 11309;
 Best Local Similarity 100.0%; Pred. No. 8,2e-213;
 Matches 966; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 AGTAACCTTATCTCAAGAGTAGACATGAAGAAAAATTTATGTCAGTGGCATCACA 60
DB 7632 AGTAACCTTATCTCAAGAGTAGACATGAAGAAAAATTTATGTCAGTGGCATCACA 7691
OY 61 CATTTATCAGTAGCAACTTTAGCAGCTTGTGAAAAGGTGTCAGAAAGTGCAGACCTTATC 120
DB 7692 CATTTATCAGTAGCAACTTTAGCAGCTTGTGAAAAGGTGTCAGAAAGTGCAGACCTTATC 7751
OY 121 AGCATGAAGGGGAGTGTCTTACAGACATCAATTTTATGAGCAAGTGAAGCAACCT 180
DB 7752 AGCATGAAGGGGAGTGTCTTACAGACATCAATTTTATGAGCAAGTGAAGCAACCT 7811
OY 181 TCAGCCCAACAAGTCTGTTAATATGACATCCAAAAAGTTTGAAGAACATATGCG 240
DB 7812 TCAGCCCAACAAGTCTGTTAATATGACATCCAAAAAGTTTGAAGAACATATGCG 7871
OY 241 TCAGAGCTTGATTAAGAGAGTGTGATGATCTATTGCGAAGAAAAAACATATGCG 300
DB 7872 TCAGAGCTTGATTAAGAGAGTGTGATGATCTATTGCGAAGAAAAAACATATGCG 7931
OY 301 GAAAGCTTACCAACGCTGCTTCACAGACAGTATGATCTCTGAAAACAGTAAAGCTCAA 360
DB 7932 GAAAGCTTACCAACGCTGCTTCACAGACAGTATGATCTCTGAAAACAGTAAAGCTCAA 7991
OY 361 ATTCGTACAGAGTAATTAAGTGGCAGTTAAGAGTGAAGAGTGAAGTGAATGACA 420
DB 7992 ATTCGTACAGAGTAATTAAGTGGCAGTTAAGAGTGAAGAGTGAAGTGAATGACA 8051
OY 421 GATGAAGCCTATTAAGAAAGCCTTTGATGATGATCACTCCAGATGTAACGGCTCAATCATC 480
DB 8052 GATGAAGCCTATTAAGAAAGCCTTTGATGATGATCACTCCAGATGTAACGGCTCAATCATC 8111
OY 481 GCTCTTAATTAATGAAGATTAAGGCCAAGAAAGTTCTGAAAAAGCCAGAGAGTGC 540
DB 8112 GCTCTTAATTAATGAAGATTAAGGCCAAGAAAGTTCTGAAAAAGCCAGAGAGTGC 8171
OY 541 GATTTTGTCTCAATTAAGCCAAAGATTAATTCAGTATGATGAAGAAAAAATGTTGGA 600
DB 8172 GATTTTGTCTCAATTAAGCCAAAGATTAATTCAGTATGATGAAGAAAAAATGTTGGA 8231
OY 601 GAAATTTACCTTTGATTTCTGCTTCACAGAGTACTGAGCAAGTCAAAAAAGCCGCTTTC 660
DB 8232 GAAATTTACCTTTGATTTCTGCTTCACAGAGTACTGAGCAAGTCAAAAAAGCCGCTTTC 8291
OY 661 GCTTTAGATGTGAGTGGTGTCTTCTGATGATTAAGCAAGTGGCACAAGCTTCACT 720
DB 8292 GCTTTAGATGTGAGTGGTGTCTTCTGATGATTAAGCAAGTGGCACAAGCTTCACT 8351
OY 721 AGCAATATTAATTAATTAAGTCACTAAGAAAAAGAAAAATCATATATATGATGAC 780
DB 8352 AGCAATATTAATTAATTAAGTCACTAAGAAAAAGAAAAATCATATATATGATGAC 8411
OY 781 TACAAAAGAAAAATTAAGAACTGTTATCTTGAAGTCAAAAAAATGATTCACATTTGTT 840
DB 8412 TACAAAAGAAAAATTAAGAACTGTTATCTTGAAGTCAAAAAAATGATTCACATTTGTT 8471
OY 841 CAAAGATTAATTAAGAAAAATTAAGCAAGCAATATCAAGGTTAAGCAAGCTTTC 900
DB 8472 CAAAGATTAATTAAGAAAAATTAAGCAAGCAATATCAAGGTTAAGCAAGCTTTC 900
OY 901 CAAATATCTTTTACCAATATATCGGTGATGAGATTAAGCTCAAGCAAGTATGATCA 960
DB 8532 CAAATATCTTTTACCAATATATCGGTGATGAGATTAAGCTCAAGCAAGTATGATCA 8591

```

OY 961 AACGAA 966
 DB 8592 AACGAA 8597

RESULT 3*
 ABS56454
 ID ABS56454 standard; DNA; 2162598 BP.

AC ABS56454;
 XX
 DT 10-FEB-2003 (first entry)

XX Streptococcus pneumoniae type 4 strain complete genome.

DE ds: bacterial meningitis; pneumonia; sepsis; otitis media; genome;
 KW ear infection; antilaminatory; antibacterial; immunostimulant;
 KW auditory; respiratory; gene therapy; vaccine.

XX Streptococcus pneumoniae type 4 strain.

XX MO200277021-A2.

XX 03-OCT-2002.

XX 27-MAR-2002; 2002MO-1B02163.

XX 27-MAR-2001; 2001GB-0007658.

XX (CHIR-) CHIRON SPA.

XX (GENO-) INST GENOMIC RES.

XX Masignani V, Tettelin H, Fraser C;

XX MPI; 2003-040579/03.

XX New proteins and nucleic acid molecules from *Streptococcus pneumoniae*,
 PT useful as medicaments for treating or preventing a disease or infection
 PT due to *Streptococcus bacteria*, such as pneumonia, sepsis, otitis media
 PT or ear infection

PS Claim 17; SEQ ID No 4979; 56bp; English.

XX The invention relates to a protein comprising or having at least 50%
 CC identity to any of the 2469 amino acid sequences, identified in the
 CC specification (available on a computer readable format), or its fragment,
 CC expressed from 2469 of 2489 identified DNA coding regions from the
 CC *Streptococcus pneumoniae* type 4 strain genomic sequence appearing as
 CC ABS56454. Also included are an antibody which binds one of the
 CC proteins, treating a patient by administering the protein, DNA or
 CC antibody (in a composition), a kit comprising first and second primers,
 CC which are the nucleic acid cited above or fragments between nucleotides
 CC 8-100 of a sequence not defined in the specification, for amplifying a
 CC target sequence contained within a *Streptococcus nucleic acid* sequence,
 CC where the first primer is substantially complementary to the target
 CC sequence and the second primer is substantially complementary to the
 CC complement of the target sequence, and where the parts of the primers
 CC having substantial complementarity define the termini of the target
 CC sequence to be amplified, assay comprising contacting a test compound
 CC with the protein, and determining whether the test compound binds to the
 CC protein and a *Streptococcus pneumoniae bacterium*, where one or more
 CC genes encoding the proteins has been rendered inactive. The proteins,
 CC nucleic acid molecules, antibody and compositions are useful as
 CC medicaments for treating or preventing a disease or infection due to
 CC *Streptococcus bacteria*, particularly *S. pneumoniae*, such as pneumonia,
 CC sepsis, otitis media or ear infection. They are also useful in developing
 CC vaccines, diagnostics and antibiotics. The methods are useful for
 CC identifying immunodominant proteins. The present sequence is the
 CC *Streptococcus pneumoniae* type 4 strain genome sequence.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

```

XX Sequence 2162598 BP; 654373 A; 427176 C; 431369 G; 649680 T; 0 other;
SQ
Query Match      100.0%; Score 966; DB 25; Length 2162598;
Best Local Similarity 100.0%; Pred No. 3.4e-212;
Matches 966; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGTAACACTATCTCAAGAGTAGACATGAGAAAAAATTTATGGCAGGTGCATCACA 60
DB 820773 AGTAACACTATCTCAAGAGTAGACATGAGAAAAAATTTATGGCAGGTGCATCACA 820832
QY 61 CATTATTCAGTAGCAACTTATAGCAGCTTGTGGAAGGGTCAGAGAGTGCAGACCTTATC 120
DB 820833 CATTATTCAGTAGCAACTTATAGCAGCTTGTGGAAGGGTCAGAGAGTGCAGACCTTATC 820892
QY 121 AGCATGAAGGGGATGTCATTACAGAAACATCAATTTATGAGCAAGTGAAGCAACCCCT 180
DB 820893 AGCATGAAGGGGATGTCATTACAGAAACATCAATTTATGAGCAAGTGAAGCAACCCCT 820952
QY 181 TCAGCCCAACAAGTCTTGTAAATATGACCATCCAAAAAGTTTGTAAAAACAATATGCG 240
DB 820953 TCAGCCCAACAAGTCTTGTAAATATGACCATCCAAAAAGTTTGTAAAAACAATATGCG 821012
QY 241 TCAGAGCTTGATGATTAAGAGTGTGATGATGATGATGATGATGATGATGATGATGATG 300
DB 821013 TCAGAGCTTGATGATTAAGAGTGTGATGATGATGATGATGATGATGATGATGATGATG 821072
QY 301 GAAACTACCAACGCTGCTTGTGCACAGCAGTATGATGATGATGATGATGATGATGATGATG 360
DB 821073 GAAACTACCAACGCTGCTTGTGCACAGCAGTATGATGATGATGATGATGATGATGATGATG 821132
QY 361 ATTGCTACAGTAATTAATAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 420
DB 821133 ATTGCTACAGTAATTAATAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 821192
QY 421 GATGAGCCTATATAGAAAGCCTTGTGATGATGATGATGATGATGATGATGATGATGATGATG 480
DB 821193 GATGAGCCTATATAGAAAGCCTTGTGATGATGATGATGATGATGATGATGATGATGATGATG 821252
QY 481 CGTCTTAATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
DB 821253 CGTCTTAATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 821312
QY 541 GATTTGGCTCAATTAACCAAGATATATGATGATGATGATGATGATGATGATGATGATGATGATG 600
DB 821313 GATTTGGCTCAATTAACCAAGATATATGATGATGATGATGATGATGATGATGATGATGATGATG 821372
QY 601 GAAATTAACCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
DB 821373 GAAATTAACCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 821432
QY 661 GCTTTAGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720
DB 821433 GCTTTAGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 821492
QY 721 AGCCAATATTAATCTTAAACTCACTAAGAAAAAGAAAAATCAATCTAATATGATGATGATGATG 780
DB 821493 AGCCAATATTAATCTTAAACTCACTAAGAAAAAGAAAAATCAATCTAATATGATGATGATGATG 821552
QY 781 TACAAAGAAAAAATTAATAAAGCTGATCTGACTCAAAAACAAAATATTCACATTTGTT 840
DB 821553 TACAAAGAAAAAATTAATAAAGCTGATCTGACTCAAAAACAAAATATTCACATTTGTT 821612
QY 841 CAAGCATTTATCGAAAAAGATTTGCAAGCAAGCCAAATATCAAGGTTAAGCAAGCAAGCTTC 900
DB 821613 CAAGCATTTATCGAAAAAGATTTGCAAGCAAGCCAAATATCAAGGTTAAGCAAGCTTC 821672
QY 901 CAAAAATATCTTTACCAATATATCGGTGTGAGATTCAGAGCTCAAGCAGTATATCATCA 960
DB 821673 CAAAAATATCTTTACCAATATATCGGTGTGAGATTCAGAGCTCAAGCAGTATATCATCA 821732
QY 961 AACGAA 966
|||||

```

```

DB 821733 AACGAA 821738
RESULT 4
ABX06679
ID ABX06679 standard; DNA; 939 BP.
XX
AC ABX06679;
XX
DT 11-FEB-2003 (first entry)
XX
DE S. pneumoniae type 4 strain coding region #967.
XX
KW Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;
KW ear infection; antiinflammatory; antibacterial; immunostimulant;
KW auditory; respiratory; gene therapy; vaccine.
XX
OS Streptococcus pneumoniae type 4 strain.
XX
PN MO20027021-A2.
XX
PD 03-OCT-2002.
XX
PF 27-MAR-2002; 2002MO-IB02163.
XX
PR 27-MAR-2001; 2001GB-0007658.
XX
PA (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
PI Masignani V, Tettelin H, Fraser C;
XX
DR WPI; 2003-040579/03.
XX
DR P-PSDB; AB001392.
XX
PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT useful as medicaments for treating or preventing a disease or infection
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
PT or ear infection
XX
XX
XX Claim 6; SEQ ID NO 1933; 56pp; English.
XX
CC The invention relates to a protein comprising or having at least 50%
CC identity to any of the 2469 amino acid sequences, identified in the
CC specification (available on a computer readable format), or its fragment,
CC expressed from 2469 of 2489 identified DNA coding regions from the
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC ABS56454. Also included are an antibody which binds one of the
CC proteins, treating a patient by administering the protein, DNA or
CC antibody (in a composition), a kit comprising first and second primers,
CC which are the nucleic acid cited above or fragments between nucleotides
CC 8-100 of a sequence not defined in the specification, for amplifying a
CC target sequence contained within a Streptococcus nucleic acid sequence,
CC where the first primer is substantially complementary to the target
CC sequence and the second primer is substantially complementary to the
CC complement of the target sequence, and where the parts of the primers
CC having substantial complementarity define the termini of the target
CC sequence to be amplified, assay comprising contacting a test compound
CC with the protein, and determining whether the test compound binds to the
CC protein and a Streptococcus pneumoniae bacterium, where one or more
CC genes encoding the proteins has been rendered inactive. The proteins,
CC nucleic acid molecules, antibody and compositions are useful as
CC medicaments for treating or preventing a disease or infection due to
CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
CC sepsis, otitis media or ear infection. They are also useful in developing
CC vaccines, diagnostics and antibiotics. The methods are useful for
CC identifying immunodominant proteins. The present sequence is one of
CC the 2489 identified coding region from the genomic sequence.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX

```

Sequence 939 BP; 363 A; 165 C; 182 G; 229 T; 0 other;

Query Match 97.2%; Score 939; DB 25; Length 939;
Best Local Similarity 100.0%; Pred. No. 6.9e-207;
Matches 939; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 28 ATGAGAAAAAATTTTGGCAGAGTCCATACATATTTTCAGTATGACCACTTTAGCAGCT 87
DB 1 ATGAGAAAAAATTTTGGCAGAGTCCATACATATTTTCAGTATGACCACTTTAGCAGCT 60
OY 88 TGTTCGAAAAGGCTCAGAAAGTGCAGACCTTATCAGCATGAAAGGGAGTCTATTACAGAA 147
DB 61 TGTTCGAAAAGGCTCAGAAAGTGCAGACCTTATCAGCATGAAAGGGAGTCTATTACAGAA 120
OY 148 CATCAATTTTATGACCAAGTGAAGCAACCTTACGCCCAAGCTCTGTTAAATATAG 207
DB 121 CATCAATTTTATGACCAAGTGAAGCAACCTTACGCCCAAGCTCTGTTAAATATAG 180
OY 208 ACCATCCAAAAAGTTTGTGAAAAACAATATGGCTCAGAGCTTGATGATTAAGAGTTGAT 267
DB 181 ACCATCCAAAAAGTTTGTGAAAAACAATATGGCTCAGAGCTTGATGATTAAGAGTTGAT 240
OY 268 GATCTATTGTCGCGAAGAAAAACAATATGGCGAAACTACCAAGCTGCTTGTGCACAA 327
DB 241 GATCTATTGTCGCGAAGAAAAACAATATGGCGAAACTACCAAGCTGCTTGTGCACAA 300
OY 328 GCAGGTATGACTCTTGAACACGTAAGCTCAATTCGTACAGTAAGTAATTAGTGTAGTGG 387
DB 301 GCAGGTATGACTCTTGAACACGTAAGCTCAATTCGTACAGTAAGTAATTAGTGTAGTGG 360
OY 388 GCAGTTAAGAAAGTAGCAGAAAGCTGATGACAGATGAAGCCTATTAAGAAAGCCTTTGAT 447
DB 361 GCAGTTAAGAAAGTAGCAGAAAGCTGATGACAGATGAAGCCTATTAAGAAAGCCTTTGAT 420
OY 448 GAGTACACTCCGAGATGTAAGGCTCAAAATCATCCGCTTATTAATGAAGATTAAGGCCAA 507
DB 421 GAGTACACTCCGAGATGTAAGGCTCAAAATCATCCGCTTATTAATGAAGATTAAGGCCAA 480
OY 508 GAGTCTTCGAAAAGGCCAAGCAGAAAGTGTGATTTGCTCAATTAAGCCAAAGATAT 567
DB 481 GAGTCTTCGAAAAGGCCAAGCAGAAAGTGTGATTTGCTCAATTAAGCCAAAGATAT 540
OY 568 TCACTGATGAAAAAACAAGAAATGCTGAGAAATTAACCTTGTATTCGCTTCACAA 627
DB 541 TCACTGATGAAAAAACAAGAAATGCTGAGAAATTAACCTTGTATTCGCTTCACAA 600
OY 628 GAAGTACCTGAGCAAGTCAAAAAGCCGCTTCGCTTGAAGTGTGATGCTGTTGTGAT 687
DB 601 GAAGTACCTGAGCAAGTCAAAAAGCCGCTTCGCTTGAAGTGTGATGCTGTTGTGAT 660
OY 688 GTGATTACAGCAACTGGCAGACAGCCTACAGTACAGCAATATTAACCTGTAACCTCAGT 747
DB 661 GTGATTACAGCAACTGGCAGACAGCCTACAGTACAGCAATATTAACCTGTAACCTCAGT 720
OY 748 AAGAAAAACAGAAAAATCATCTAATATGATGACTACAAAGAAAAATTTAAAACTGTATC 807
DB 721 AAGAAAAACAGAAAAATCATCTAATATGATGACTACAAAGAAAAATTTAAAACTGTATC 780
OY 808 TTGACTCAAAAACAAATGATTCACATCTGTTCAAGCATATTCGGAAGAAAGATTGCAA 867
DB 781 TTGACTCAAAAACAAATGATTCACATCTGTTCAAGCATATTCGGAAGAAAGATTGCAA 840
OY 868 GCAGCAATATATCAAGTTAAGAGCAAGCCTTCGAAATATCTTTACCAATATATACGCT 927
DB 841 GCAGCAATATATCAAGTTAAGAGCAAGCCTTCGAAATATCTTTACCAATATATACGCT 900
OY 928 GGTGAGATTCAAGCTCAAGCAGTACTACATCAAGAA 966
DB 901 GGTGAGATTCAAGCTCAAGCAGTACTACATCAAGAA 939

```

RESULT 5
AAA05430

```

ID AAA05430 standard; DNA; 942 BP.
XX
AC AAA05430;
XX
XX 24-MAY-2000 (first entry)
DT
DE Streptococcus pneumoniae nucleotide sequence ID212 - 4136.1.
XX
KW Streptococcus pneumoniae; vaccine; screening; protein antigen;
KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;
KW pneumococcal disease; ds.
OS Streptococcus pneumoniae.
XX
PN MO200006737-A2.
XX
PD 10-FEB-2000.
XX
XX 27-JUL-1999; 99WO-GB02451.
PF
PR 27-JUL-1998; 98GB-0016337.
PR 19-MAR-1999; 99US-0125164.
XX
PA (MICR-) MICROBIAL TECHNIQS LTD.
XX
PI Gilbert CFG, Hansbro PM;
XX
DR WPI: 2000-195300/17.
DR P-PSDB; AA81675.
XX
XX New Streptococcal protein, useful as a vaccine, for diagnosis of
PT pneumococcal diseases and for screening agents capable of antagonizing
PT or inhibiting expression of the protein
XX
PS Claim 6; Page 92; 108pp; English.
XX
XX AA81501 to AA81679 represent specifically claimed protein sequences
CC isolated from Streptococcus pneumoniae. AA05407 to AA05590 represent
CC specifically claimed nucleotide sequences isolated from S. pneumoniae.
CC The sequences have antibacterial and antiinflammatory properties.
CC The protein sequences, and fragments of them, are useful as immunogens
CC and/or antigens. The nucleotide sequences can be used in vaccines and in
CC diagnostic assays. The proteins and nucleotides can be useful for the
CC detection and diagnosis of S. pneumoniae. The protein sequences are also
CC useful for screening an agent capable of antagonizing, inhibiting or
CC interfering with the function or expression of the proteins in which the
CC agent is useful for treatment or prophylaxis of S. pneumoniae infection
CC and meningitis. AA05591 to AA05614 represent primers used in the
CC exemplification of the present invention.
XX
SQ Sequence 942 BP; 364 A; 165 C; 183 G; 230 T; 0 other;

```

Query Match 97.2%; Score 939; DB 21; Length 942;
Best Local Similarity 100.0%; Pred. No. 6.9e-207;
Matches 939; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 28 ATGAGAAAAAATTTTGGCAGAGTCCATACATATTTTCAGTATGACCACTTTAGCAGCT 87
DB 1 ATGAGAAAAAATTTTGGCAGAGTCCATACATATTTTCAGTATGACCACTTTAGCAGCT 60
OY 88 TGTTCGAAAAGGCTCAGAAAGTGCAGACCTTATCAGCATGAAAGGGAGTCTATTACAGAA 147
DB 61 TGTTCGAAAAGGCTCAGAAAGTGCAGACCTTATCAGCATGAAAGGGAGTCTATTACAGAA 120
OY 148 CATCAATTTTATGACCAAGTGAAGCAACCTTACGCCCAAGCTCTGTTAAATATAG 207
DB 121 CATCAATTTTATGACCAAGTGAAGCAACCTTACGCCCAAGCTCTGTTAAATATAG 180
OY 208 ACCATCCAAAAAGTTTGTGAAAAACAATATGGCTCAGAGCTTGATGATTAAGAGTTGAT 267
DB 181 ACCATCCAAAAAGTTTGTGAAAAACAATATGGCTCAGAGCTTGATGATTAAGAGTTGAT 240
OY 268 GATCTATTGTCGCGAAGAAAAACAATATGCGGAAAGTACCAAGCTGCTTGTGCACAA 327

```

```

DB      241 GATATATTGGCGAAGAAAAAACAATATGGCGAAACTACCAACGTCTTGCACAA 300
      328 GCAGTATGACTCTTGAACACGTAAAGCTCAATTCGTACAGTAATTAAGTGTG 387
      301 GCAGTATGACTCTTGAACACGTAAAGCTCAATTCGTACAGTAATTAAGTGTG 360
      388 GCAGTATGAGGAGTGAAGCTGATTTGACAGTGAAGCTTAAAGAACCTTTGAT 447
      361 GCAGTATGAGGAGTGAAGCTGATTTGACAGTGAAGCTTAAAGAACCTTTGAT 420
      448 GAGTACTCCAGATGTAAAGCTCAATTCGTCTTAATTAATGAAGTAAGGCCAA 507
      421 GAGTACTCCAGATGTAAAGCTCAATTCGTCTTAATTAATGAAGTAAGGCCAA 480
      508 GAGTCTTCGAAAAAGCCAGGACAGAGTGTCTGATTTTCCATTTGCGCAAGATAT 567
      481 GAGTCTTCGAAAAAGCCAGGACAGAGTGTCTGATTTTCCATTTGCGCAAGATAT 540
      568 TCACTGATGAAAAAACAAGAAATGGTGAAGAAATTAACCTTGTCTTCAACA 627
      541 TCACTGATGAAAAAACAAGAAATGGTGAAGAAATTAACCTTGTCTTCAACA 600
      628 GAGTACTGAGCAAGTCAAAAAGCCCTTTCGTTTGAATGTGAGTGTCTTCTGAT 687
      601 GAGTACTGAGCAAGTCAAAAAGCCCTTTCGTTTGAATGTGAGTGTCTTCTGAT 660
      688 GAGTACTGAGCAAGTCAAAAAGCCCTTTCGTTTGAATGTGAGTGTCTTCTGAT 747
      661 GAGTACTGAGCAAGTCAAAAAGCCCTTTCGTTTGAATGTGAGTGTCTTCTGAT 720
      748 AAGAAAAAGAAAAATCATCTAATATGTGACTACAAAGAAAAATTAAGAACTGTATC 807
      721 AAGAAAAAGAAAAATCATCTAATATGTGACTACAAAGAAAAATTAAGAACTGTATC 780
      808 TTGACTCAAAAAACAATATGATTCACATTTGTTCAAGCATTTTCGGAAGAAATTCGA 867
      781 TTGACTCAAAAAACAATATGATTCACATTTGTTCAAGCATTTTCGGAAGAAATTCGA 840
      868 GAGCAATATCAAGGTTAAGACCAAGCCCTTCCAAATATCTTTACCAATATATCGST 927
      841 GAGCAATATCAAGGTTAAGACCAAGCCCTTCCAAATATCTTTACCAATATATCGST 900
      928 GGTGAGATTCAGCTCAAGCAAGTGTACATCAACGAA 966
      901 GGTGAGATTCAGCTCAAGCAAGTGTACATCAACGAA 939
  
```

RESULT 6
AAV27340
ID AAV27340 standard; DNA; 640 BP.
XX
AC AAV27340;
XX
DT 02-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae SP0021 nucleotide.
XX
KM Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
XX detection; pneumonia; otitis media; meningitis; ss.
OS Streptococcus pneumoniae.
XX
XX key Location/Qualifiers
XX CDS 2..640
XX FT /*tag= a
XX FT /product= "SP0021"
XX FT /note= "no stop codon given"
XX PD NO9818930-A2.
XX 07-MAY-1998.

```

PF 30-OCT-1997; 97WO-US19422.
XX
XX 31-OCT-1996; 96US-0029960.
XX
XX (HOMA-) HUMAN GENOME SCI INC.
XX
XX Chai GH, Hromockyj A, Johnson LS, Kunsch CA.
XX
XX WPI: 1998-27224/24.
XX P-PSDB: AAM55079.
XX
XX Nucleic acid encoding antigenic peptide(s) from Streptococcus
XX pneumoniae - or their epitope-containing fragments, useful in
XX protective or therapeutic vaccines, and for diagnosis
XX
PS Claim 1; Page 55; 118pp; English.
XX
XX The present sequence encodes a protein from Streptococcus pneumoniae.
XX The nucleic acid sequence encoding the Streptococcus pneumoniae protein
XX can be useful in vaccines for inducing protective antibodies against
XX Streptococcus pneumoniae, for treatment or prevention of infection e.g.
XX pneumonia, otitis media or meningitis. Probes based on the nucleic acid
XX are used to detect Streptococcus infection (by usual hybridisation or
XX amplification methods), also for isolating Streptococcus genes or their
XX allelic variants. The protein can be used similarly to detect specific
XX antibodies in standard immunoassays, especially for diagnosing or
XX monitoring infections. Antibodies which bind the protein are used to
XX detect corresponding antigens, to purify the protein and for passive
XX immunisation (optionally coupled to a toxin). Vaccines are administered,
XX e.g. by injection, orally or through the skin, typically at 0.01-1000
XX (especially 10-300) mu g/ml per dose.
XX
SQ Sequence 640 BP; 234 A; 112 C; 141 G; 153 T; 0 other;
XX
XX Query Match 57.2%; Score 552.4; DB 19; Length 640;
XX Best Local Similarity 94.7%; Pred. No. 8.2e-118;
XX Matches 595; Conservative 0; Mismatches 26; Indels 7; Gaps 2;
  
```

90 TTCGAAGGGTCAGAAAGTGCAGACCTTATCGCATGAAAGGGATGTCATTACGAACA 149
1 TTCGAAGGGTCAGAAAGTGCAGACCTTATCGCATGAAAGGGATGTCATTACGAACA 60
150 TCATTTTATGAGCAAGTGAAGCAACCCCTTCAGCCCAACAGTCTTGAATATGAC 209
61 TCATTTTATGAGCAAGTGAAGCAACCCCTTCAGCCCAACAGTCTTGAATATGAC 120
210 CATCCAAAAGTTTGAAGAAACAATATGCTCAGAGCTTGATGATTAAGAGCTTGATGA 269
121 CATCCAAAAGTTTGAAGAAACAATATGCTCAGAGCTTGATGATTAAGAGCTTGATGA 180
270 TACTATGCGGGAAGAAAAAACAATATGCGGAAGAAATCAACAGTGTCTTGCACAAGC 329
181 TACTATGCGGGAAGAAAAAACAATATGCGGAAGAAATCAACAGTGTCTTGCACAAGC 240
330 AGGTATGACTCTTGAACACAGTAAAGCTCAATTCGTACAACTAATTAAGTGAAGTGGC 389
241 AGGTATGACTCTTGAACACAGTAAAGCTCAATTCGTACAACTAATTAAGTGAAGTGGC 300
390 AGTTAAGAAAGTGAAGCAAGAGTGAATTCACATGAAGCTTAAAGAAAGCTTTGATGA 449
301 AGTTAAGAAAGTGAAGCAAGAGTGAATTCACATGAAGCTTAAAGAAAGCTTTGATGA 360
450 GTACACTCCAGATGTAAAGCTCAATTCGTCTTAATTAATGAAGTAAGGCCAAAGA 509
361 GTACACTCCAGATGTAAAGCTCAATTCGTCTTAATTAATGAAGTAAGGCCAAAGA 420
510 AGTTCTCGAAAAAGCCAGAGAGTGTCTGATTTTGCCTCAATTAGCCAAAGATAATTC 569
421 AGTTCTCGAAAAAGCCAGAGAGTGTCTGATTTTGCCTCAATTAGCCAAAGATAATTC 480
570 AACTGATGAAAAAACAAGAAATGGTGAAGAAATTAACCTTGAATTCGCTTCAACAGA 629
481 AACTGATGAAAAAACAAGAAATGGTGAAGAAATTAACCTTGAATTCGCTTCAACAGA 540

QY 630 AGTACCTG--AGCAAGTCAAAAAAGCCGCTTTCGCTTAGATGATGATGCTTCTGAT 687
 DB 541 AGTACCTGAGCAAGTCCAAAAAGCCGCTTTCGCTTAGATGATGATGCTTCTGAT 600
 QY 688 G-----TGATTACAGCAACTGGCAGACA 710
 DB 601 GGATGTGGATTACAGCAACTGGCAGACA 628

RESULT 7

AB084808
 ID AB084808 standard; DNA; 640 BP.

AC AB084808;

DT 04-SEP-2002 (first entry)

DE S. pneumoniae SP021 nucleotide sequence SEQ ID NO:33.

KM Streptococcus pneumoniae: epitope; vaccine; antigenic protein;
 KM antibacterial; Streptococcal infection; detection; gene; ds.

OS Streptococcus pneumoniae.

PN US2002061545-A1.

PD 23-MAY-2002.

PF 22-JAN-2001; 2001US-0765272.

PR 30-OCT-1997; 97US-0961083.

XX (CHOI/) CHOI G H.
 PA (KUNS/) KUNSCH C A.
 PA (BARA/) BARASH S C.
 PA (DILL/) DILLON P J.
 PA (DOUG/) DOUGHERTY B.
 PA (FANN/) FANNON M R.
 PA (ROSE/) ROSEN C A.

PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
 PI Rosen CA;

DR WPI; 2002-479261/51.
 DR P-PSDB; ABP54573.

XX New Streptococcus pneumoniae antigens, useful for detecting
 PT Streptococcus and for preventing or attenuating disease caused by
 PT Streptococcus infection -

PS Claim 1; Page 24; 70pp; English.

CC AB084792 to AB084904 represents nucleic acids which encode the
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.
 CC The S. pneumoniae antigens have antibacterial activity and can be
 CC used in vaccines. The S. pneumoniae antigens can also be used to
 CC prevent or attenuate a Streptococcal infection in an animal. The
 CC polynucleotides encoding the S. pneumoniae antigens can be used to
 CC detect Streptococcus nucleic acids. AB084905 to AB085130 represent
 CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)
 CC which are used in an example from the present invention.

XX Sequence 640 BP; 234 A; 112 C; 141 G; 153 T; 0 other;

Query Match 57.2%; Score 552.4; DB 24; Length 640;

Best Local Similarity 94.7%; Pred.No. 8.2e-118;
 Matches 595; Conservative 0; Mismatches 26; Indels 7; Gaps 2;

QY 90 TTCGAAAGGTCAGAAAGTGCAGACTTATCAGCATGAAGGGATGTCATTACGAGACA 149
 DB 1 TTCGAAAGGTCAGAAAGTGCAGACTTATCAGCATGAAGGGATGTCATTACGAGACA 60

QY 150 TCAATTTTATGAGCAAGTGAAGCAACCCCTTACGCCCAACAGTCTTGTAAATATGAC 209
 DB 61 TCAATTTTATGAGCAAGTGAAGCAACCCCTTACGCCCAACAGTCTTGTAAATATGAC 120
 QY 210 CATCCAAAAAGTTTTGAAAAACAATATGCGTACAGCTTGTATGAAGAGCTTGATGA 269
 DB 121 CATCCAAAAAGTTTTGAAAAACAATATGCGTACAGCTTGTATGAAGAGCTTGATGA 180
 QY 270 TACTATTTGCGGAGAAAAAACAATATGCGGAGAAAAACAATATGCGTACAGCTTGTATGA 329
 DB 181 TACTATTTGCGGAGAAAAAACAATATGCGGAGAAAAACAATATGCGTACAGCTTGTATGA 240
 QY 330 AGGTATGACTCTTGAAGACGTAAGCTCAATTTGTACAGTAATATGATGAGTGGC 389
 DB 241 AGGTATGACTCTTGAAGACGTAAGCTCAATTTGTACAGTAATATGATGAGTGGC 300
 QY 390 AGTTAAGAAAGTACGAGAAAGCTGAATTTGACAGATGAGAGCTTGAAGAAAGCTTTGATGA 449
 DB 301 AGTTAAGAAAGTACGAGAAAGCTGAATTTGACAGATGAGAGCTTGAAGAAAGCTTTGATGA 360
 QY 450 GTACACTCCAGATGTACAGGCTCAATTCGCTTAATATGAAGTAAGCCAAAGA 509
 DB 361 GTACACTCCAGATGTACAGGCTCAATTCGCTTAATATGAAGTAAGCCAAAGA 420
 QY 510 AGTCTCGAAAAAGCCAAAGCAGAGTGTGATTTTGTCTCAATTAGCCAAAGATATTC 569
 DB 421 AGTCTCGAAAAAGCCAAAGCAGAGTGTGATTTTGTCTCAATTAGCCAAAGATATTC 480
 QY 570 AACTGATGAAAAAAGCAAAAAATGCTGAGAAATTTACCTTTGATTTCTGCTTCAACAGA 629
 DB 481 AACTGATGAAAAAAGCAAAAAATGCTGAGAAATTTACCTTTGATTTCTGCTTCAACAGA 540
 QY 630 AGTACCTG--AGCAAGTCAAAAAAGCCGCTTTCGCTTAGATGATGATGCTTCTGAT 687
 DB 541 AGTACCTGAGCAAGTCCAAAAAGCCGCTTTCGCTTAGATGATGATGCTTCTGAT 600
 QY 688 G-----TGATTACAGCAACTGGCAGACA 710
 DB 601 GGATGTGGATTACAGCAACTGGCAGACA 628

RESULT 8

AA30925
 ID AAX30925 standard; DNA; 451 BP.

AC AAX30925;

DT 20-MAY-1999 (first entry)

DE Streptococcus pneumoniae genomic DNA sequence SEQ ID NO:202.

KM Streptococcus pneumoniae strain 0100993; vaccine; immune response;
 KM Streptococcal infection; pneumococcal; ss.

OS Streptococcus pneumoniae.

PN MO9737026-A1.

PD 09-OCT-1997.

PF 01-APR-1997; 97WO-US05306.

PR 22-AUG-1996; 96US-0025788.

PR 02-APR-1996; 96US-0014690.

PA (SMIR) SMITHKLINE BEECHAM CORP.
 PA (SMIR) SMITHKLINE BEECHAM PLC.

PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;
 PI Stoolia RK;

DR WPI; 1997-50311/46.
 DR P-PSDB; AAY11343.

XX Nucleic acids encoding pneumococcal polypeptide(s) - useful in
PT vaccines, drug screening, etc
XX

PS Claim 5: Page 172: 354pp: English.

XX AAX30724 to AAX30946 represent genomic DNA sequences isolated from
CC Streptococcus pneumoniae strain 0100993. These genomic DNA sequences
CC encode the novel proteins given in AAY1114 to AAY1137. The proteins,
CC isolated from Streptococcus pneumoniae, can be used in vaccines against
CC streptococcal infections and in assays for identifying compounds that
CC inhibit or activate the activity of the proteins. The antagonists can
CC be used to treat an individual having need to inhibit a bacterial
CC protein. Vectors expressing the proteins can be used to induce a
CC protective immune response in mammals.

XX Sequence 451 BP: 170 A; 77 C; 87 G; 116 T; 1 other;

Query Match 27.5%; Score 265.6; DB 19; Length 451;
Best Local Similarity 97.9%; Pred. No. 9.7e-52;
Matches 279; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

OY 1 AGTAACACTTATCTCAAGAGTAGACATGAGAAAAAATTATGCGACGTCATCACA 60
DB 167 AGTAACACTTATCTCAAGAGTAGACATGAGAAAAAATTATGCGACGTCATCACA 226
OY 61 CTATTATCAGTAGCAACTTTAGCAGCTTGTGGAAGGGTCAGAAAGTCCAGACCTTATC 120
DB 227 CTATTATCAGTAGCAACTTTAGCAGCTTGTGGAAGGGTCAGAAAGTCCAGACCTTATC 286
OY 121 AGCATGAAGGGGATGCTCATTCACAGCAATTTTATGACCAAGTGAAGCAACCT 180
DB 287 AGCATGAAGGGGATGCTCATTCACAGCAATTTTATGACCAAGTGAAGCAACCT 346
OY 181 TCAGCCCAACAAGCTTGTGTAATATGACATCCAAAAAGTTTGG-AAAAACAATATG 239
DB 347 TCAGCCCAACAAGCTTGTGTAATATGACATCCAAAAAGTTTGG-AAAAACAATATG 406
OY 240 CTCAGAGCTTGATGATTAAGAGCTTGATGATCTATTCGCCGAGA 284
DB 407 CTCAGAGCTTGATGATTAAGAGCTTGATGATCTATTCGCCGAGA 451

RESULT 9

AAT98753
ID AAT98753 standard: DNA; 791 BP.

AC AAT98753;

DE 10-NOV-1998 (first entry)

XX DNA encoding a S. pneumoniae protein of unknown function.

XX Streptococcus pneumoniae protein; genetic immunisation; antagonist;
KW immunological response; inoculation; antibody production; inhibitor;
KW T cell immune response; antimicrobial compound; bacterial adhesion;
KW extracellular matrix protein; protein-mediated cell invasion; wound;
KM pathogenesis; ss.

XX Streptococcus pneumoniae.

OS Streptococcus pneumoniae.
FH Key Location/Qualifiers
FT 534..758
FT CDS /*tag= a

XX WO9743303-A1.

XX 20-NOV-1997.

XX 14-MAY-1997; 97WO-US07950.

XX 14-MAY-1996; 96US-0017670.

PA (SMK) SMITHKLINE BEECHAM CORP.
PA (SMK) SMITHKLINE BEECHAM PLC.

PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;
PI Stodola RK;

DR WPI: 1998-008793/01.

DR P-PSDB; AAW38711.

PT Novel Streptococcus pneumoniae proteins and related DNA - useful for
PT diagnosing anti-microbial agents for treatment of bacterial
PT infections

PS Claim 4: Page 239; 483pp: English.

XX This sequence encodes a Streptococcus pneumoniae protein of unknown
CC function, and represents a DNA sequence of the invention.
CC The DNA sequences were isolated from Streptococcus pneumoniae strain
CC 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
CC invention can be used to identify compounds which interact with and
CC inhibit or activate the activity of the proteins. Antagonists can be
CC used to treat diseases caused by S. pneumoniae proteins, through genetic
CC immunisation. They can also be used to induce an immunological response
CC in a mammal by inoculation with the S. pneumoniae proteins or delivery
CC of the encoding nucleic acids in a vector adequate to produce antibody
CC and/or T cell immune responses to protect the animal from disease. The
CC proteins can also be used to identify antimicrobial compounds which are
CC capable of inhibiting their bioactivity. In particular the proteins of
CC the invention can be used to prevent adhesion of bacteria to mammalian
CC extracellular matrix proteins on in-dwelling devices or in wounds, to
CC block protein-mediated mammalian cell invasion, and to block the normal
CC progression of pathogenesis in infections initiated other than by the
CC implantation of in-dwelling devices or other surgical techniques.

XX Sequence 791 BP: 258 A; 140 C; 168 G; 224 T; 1 other;

Query Match 27.5%; Score 265.6; DB 19; Length 791;
Best Local Similarity 97.9%; Pred. No. 1.1e-51;
Matches 279; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

OY 1 AGTAACACTTATCTCAAGAGTAGACATGAGAAAAAATTATGCGACGTCATCACA 60
DB 507 AGTAACACTTATCTCAAGAGTAGACATGAGAAAAAATTATGCGACGTCATCACA 566
OY 61 CTATTATCAGTAGCAACTTTAGCAGCTTGTGGAAGGGTCAGAAAGTCCAGACCTTATC 120
DB 567 CTATTATCAGTAGCAACTTTAGCAGCTTGTGGAAGGGTCAGAAAGTCCAGACCTTATC 626
OY 121 AGCATGAAGGGGATGCTCATTCACAGCAATTTTATGACCAAGTGAAGCAACCT 180
DB 627 AGCATGAAGGGGATGCTCATTCACAGCAATTTTATGACCAAGTGAAGCAACCT 686
OY 181 TCAGCCCAACAAGCTTGTGTAATATGACATCCAAAAAGTTTGG-AAAAACAATATG 239
DB 687 TCAGCCCAACAAGCTTGTGTAATATGACATCCAAAAAGTTTGG-AAAAACAATATG 746
OY 240 CTCAGAGCTTGATGATTAAGAGCTTGATGATCTATTCGCCGAGA 284
DB 747 CTCAGAGCTTGATGATTAAGAGCTTGATGATCTATTCGCCGAGA 791

RESULT 10

ABN69360
ID ABN69360 standard: DNA; 1053 BP.

AC ABN69360;

DE 01-JUL-2002 (first entry)

XX Streptococcus polynucleotide SEQ ID NO 6633.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;

KW antiInflammatory; infection; vaccine; meningitis; gene therapy; ds.
 XX Streptococcus pyogenes.
 XX
 XX WO200234771-A2.
 XX
 XX 02-MAY-2002.
 XX
 XX 29-OCT-2001; 2001WO-GB04789.
 XX
 XX 27-OCT-2000; 2000GB-0026333.
 PR 24-NOV-2000; 2000GB-0028727.
 PR 07-MAR-2001; 2001GB-0005640.
 XX
 XX (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 XX Telford J, Maignani V, Margalit Ros YI, Grandi G, Fraser C;
 PI Tettein H;
 XX
 DR MPI; 2002-352536/38.
 DR P-PSDB; ABP26729.
 XX
 XX New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -
 XX
 XX Claim 7; Page 3825; 4525pp; English.
 XX
 XX The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and antiInflammatory
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 XX
 XX Sequence 1053 BP; 415 A; 176 C; 204 G; 258 T; 0 other;
 SO
 Query Match 27.1%; Score 261.6; DB 24; Length 1053;
 Best Local Similarity 56.9%; Pred. No. 1e-50;
 Matches 505; Conservative 0; Mismatches 374; Indels 9; Gaps 1;

QY 334 ATGACTCTTGAAACACGATGAAGCTCAATTCGTACAGTAAATAGTTAGCTGGCACTT 393
 DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 313 TTGACACCTGAGACTTTTAAAGGTGAGATCCGCTCTTCAAAATTTAGTAATATCGGTT 372
 QY AAGAAGGTAGCAGAACTGATTTGACACATGAAGCTTTAAGAAAGCTTTGATAGTAC 453
 DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 373 AAAGAAGCAGCTTAAAGAAATTTGACACACACAGATTTAAGAAAGCATATGAATCTAT 432
 QY ACTCCAGATGTAAAGGCTCAATCATCCGTTCTTAATTAAGATTAAGCCAAAGAGTT 513
 DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 433 ACTCCAAATATGCGAGTGAATGATTTACTTAGTAAATGAGACAGCTAAATCAATC 492
 QY CTCGAAAAAGCCAAAGCAGAAAGTGTGCTGATTTGCTCAATTAAGCCAAAGATTAATCACT 573
 DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 493 TTAGAGGAACTTAAAGCCGAAAGGCGCAGACTTTACAGTATTTGCTAAAGAAAAAACAC 551
 QY 574 GATGAAAAACAAAGAAATGTTGAGAAATTTACCTTTGATTTGCTTCAACAGACTA 633
 DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 552 -----AACACCTGAGAAAAAAAGTGAACCTTAATTAATTTGATTCAGTGCACAAATGTA 603
 QY 634 CCGAGCAAGTCAAAAAAGCCGCTTTCGTTAGATGTGATGATGATGATGATGATGAT 693
 DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 604 CCGACTGATGTGTAAAGCCGCTTCAAGTTTGAATGAGGCGCATATCAGACGTTATC 663
 QY 694 ACAGCAACTGGCAGCAAGCCCTACAGTACCAATATTAATTTAAACCTCACTAAGAA 753
 DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 664 TCGGTTTATGATCCAACTCTTATCAAAAGAAAGTTTACATTTGTAAGGTGACTAAAAA 723
 QY 754 ACAGAAAAATCATATTTATTTGATGACTACAAAGAAATTTAAACGTTATCTTGACT 813
 DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 724 GCAGAAAAAAATGATTTGCGCAAGATTAAGAAACGTTGAAAGCTATCTATTTACT 783
 QY 814 CAAAACAAATGATTTCAATTTGTTCAAAAGCTTTATGCAAAAGAAATTTGCAACGAC 873
 DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 764 GAAATATCAAAAGATTTGATTTTCCAAACAAAGTTATTTGCAAAATGATTTGATTAAGCT 843
 QY 874 AATATCAAGTTAAGACCAAGCTTCCAAATTTATTTTCAATAT 921
 DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 844 AATGAAAAATTAAGACAAAGCTTTGCTAATATTTTGGCGCAATAT 891
 RESULT 11
 ABN71434
 ID ABN71434 standard; DNA; 927 BP.
 XX
 AC ABN71434;
 XX
 XX 01-JUL-2002 (first entry)
 DT
 XX
 DE Streptococcus polynucleotide seq ID NO 10781.
 XX
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
 KW antiInflammatory; infection; vaccine; meningitis; gene therapy; ds.
 OS Streptococcus agalactiae.
 XX
 PN WO200234771-A2.
 XX
 PD 02-MAY-2002.
 XX
 XX 29-OCT-2001; 2001WO-GB04789.
 PF
 XX 27-OCT-2000; 2000GB-0026333.
 PR 24-NOV-2000; 2000GB-0028727.
 PR 07-MAR-2001; 2001GB-0005640.
 XX
 XX (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 XX Telford J, Maignani V, Margalit Ros YI, Grandi G, Fraser C;
 PI Tettein H;
 XX

DR WPI: 2002-352536/38.
 DR P-PSDB: ABP30803.
 PT New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -
 XX
 PS Claim 7: Page 4184; 4525pp; English.
 XX
 CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC Streptococcus pyogenes (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and anti-inflammatory
 CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 XX
 XX Sequence 927 BP; 370 A; 152 C; 161 G; 244 T; 0 other;
 SQ
 Query Match 26.9%; Score 259.4; DB 24; Length 927;
 Best local similarity 57.5%; Pred. No. 3.2e-50;
 Matches 511; Conservative 0; Mismatches 366; Indels 12; Gaps 2;
 QY 36 AAAATTTATTTGCGAGTGCCTTACACATTAATTCAGTACGACCTTTAGCAGCTTGTTC--- 92
 DB 15 AAAATTTGCGAGCTTTTGTGACATTAATTCAGTACGACCTTTAGCAGCTTGTTCAGG 74
 QY 93 GAAAGGTCAGAGAGGTCAGCTTATCGCATGAAAGGGATGTCATTCAGAACATCA 152
 DB 75 TAAACATCAATTAATGACAAATGTTGTCATTAATGAAAGGCGACACTATTACGTCCTGA 134
 QY 153 ATTTATGAGCAAGTAAAGCAACCTTCAGCCCAAGCAAGTCTGTTAAATATGACAT 212
 DB 135 TTTTATGATCAAGTAAACAAATGCTGACACAAACAAATGCTTACATTTGATCTCT 194
 QY 213 CCAAAAAGTTTGTGAAAAACATATGCTCAGACCTTGATGTTAAAGAGCTTGATGATAC 272
 DB 195 CTCAGCTGTTTGTATACACAGTATGCTGATTAAGTTTCAGATTAAGAAAGTATCAGAGC 254
 QY 273 TATTGCCGAGAAAAAATATATGCGGAAACTACCAACGTCCTGTCACAAGCAGG 332
 DB 255 TTATATAAGACAGCTAAAGCTATGCTATTCATTTTCAAGCGCACTTTCACAAGCAGG 314
 QY 333 TATGACTCTTGAACACGTAAGCTCAATTTGTCACAACTAAATTAATTAAGTTGAGTGCAGT 392
 DB 315 TTTGACTCGGAAAGTTACAAACAAATTCGCAACAACTATGCTGGAATATGCTGT 374
 QY 393 TAAAGAGTGAAGAGAGCTGATTAACATGACATTAACCTTTAAAGAAAGCTTTGATGAGTA 452
 DB 375 AAAAGACGACGTAAGAAAGATTAACAGAACCAACTTAAGAAAGATTAAGAAAGCTTA 434
 QY 453 CACTCCAGATGTAAAGGCTCAATATCGTCTTAATATGAAGATTAAGGCAAGAAAGT 512
 DB 435 TACTCTGAAACTTGTATCAAGTATCAATTAATGATGACAGAGATTAAGCTTAATCTGT 494
 QY 513 TCTCGAAAAAGCAAGAGAGTGTGATTTGCTCAATTAAGCAAAAGATTAATTTCAAC 572
 DB 495 CCTTAAGATGTAAAGGCTGATGAGGATTTTTCAGAAAGATTCGCAAAAGAAAAACAAC 554
 QY 573 TGATAAAAACAAAGAAATATGTTGAGAAATTAACCTTTGATTTGCTTCAACAGAAAGT 632
 DB 555 AGCTACTGATTAAGAAAGTT-----GAGTATTAATTTGATTTCTGACAGGAGCAACCT 605

QY 633 ACCTGAGCAAGTCAAAAAAGCCGCTTCGCTTAGATGTGATGTTCTGATGTGAT 692
 DB 606 CCTTAAGAAAGTATATGTCAGACGCTTTAAGCTAGATTAATTAATGTTTTCAGATGTGCT 665
 QY 693 TACGCACTGCGACCAAGACCTTACAGTACGCAATTTACATTTGTAAGAACTCAAGAA 752
 DB 666 TTCACGCTGATTCACCAACTTATTAACAGTTACTACATCATTAAGTAAGTATGATTA 725
 QY 753 AACGAAATATCATTAATATGATGATGACAGAAAGAAATTAATTAACGTTATCTTGAC 812
 DB 726 GACAGGAAAAATATCTGATTTGCAAAATCTTCAAAAAATCCGTTAAAGAAAGTATTTTAA 785
 QY 813 TCAAAAAACAAATGATTCAACATTTGTTCAAAAGCATTAATCGMAAAGAAATTCAGACAGC 872
 DB 786 GGTAAACAAAGCATGATGAGCCCTTCAAAATTAAGATTTTCAAAAGCCTTAGAAAAAGC 845
 QY 873 CATATCAAGTTAAGACCAAGCTTCCAAAATATCTTACCAATAT 921
 DB 846 TATGTTAAGATTAAGATTAAGCATTTGACAGCATCTTATCAATAT 894

RESULT 12

ABN71519
 ID ABN71519 standard; DNA; 927 BP.
 AC ABN71519;
 XX
 XX

DT 01-JUL-2002 (first entry)
 XX

DE Streptococcus polynucleotide SEQ ID NO 10951.
 XX

KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
 KW anti-inflammatory; infection; vaccine; meningitis; gene therapy; ds.
 OS Streptococcus agalactiae.

PN WO200234771-A2.
 XX

PD 02-MAY-2002.
 XX

PF 29-OCT-2001; 2001MO-GB04789.
 XX

PR 27-OCT-2000; 2000GB-0026333.
 XX

PR 24-NOV-2000; 2000GB-002827.
 XX

PR 07-MAR-2001; 2001GB-0005640.
 XX

PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX

PI Telford J, Maignani V, Margarit Ros YI, Grandi G, Fraser C;
 PI Telford H;
 XX

DR WPI: 2002-352536/38.
 XX

P-PSDB: ABP30888.
 XX

PT New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -
 XX

PS Claim 7: Page 4192; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B
 CC Streptococcus pyogenes (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and anti-inflammatory
 CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be


```

QY 153 ATTATGAGCAAGTGAAGAAACACCCCTTCAGCCCAACAAGCTTGTGTAATATGACAT 212
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778007 TTTTATGATCAATCAAAACATCAAAAGCTGACAAATCAATGCTTACATGATCCCT 778066
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 213 CCAAAAAGTTTGAAGAAACATATGCGCTCAGAGCTGATGATTAAGAGCTTGATGATAC 272
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778067 CTCAGCTGTTTGTATACACAGATGCGTATGAAGTTGATGATTAAGAGCTTGATGATAC 778126
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 273 TATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 332
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778127 TATATATAGACAGCTAAAGGCTATGATGATGATGATGATGATGATGATGATGAT 778186
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 333 TATGACTCTTGAAGACGTAAGCTCAATTTGCTACAGATTAATTTAGTTGAGTTGGCACT 392
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778187 TTTGACTCGGAGAGCTTACCAACCAATTTGCAACACTATGCTTGTGGAATATGCTGT 778246
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 393 TAAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 452
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778247 AAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 778306
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 453 CACTCCAGATGTACGGCTCAATCATCCGCTTATATATATGAAGATTAAGCCCAAGAGAGT 512
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778307 TACTCTGAGAACTCTGTACAGATCAATTAATGATGAGAGAGATTAAGCTAATCTGT 778366
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 513 TCTGCAAAAAGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 572
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778367 CCTTAAAGATGTAAAGGCTGATGAGAGCTGATTTTGCAAGAGATTGCAAAAGAGAGAGAG 778426
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 573 TGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 632
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778427 ACCTACTGATTAAGAAAGT-----GAGTATTAATTTGATTTGCGAGGGGCAACCT 778477
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 633 ACCCTGAGCAGTCAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 692
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778478 CCTTAAAGATGTAAAGGCTGATGAGAGCTTAAAGCTAATTAAGAGAGAGAGAGAGT 778537
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 693 TACAGCAACTGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 752
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778538 TTCAACGGTGTATTCACACACTTATTAACAAAGTTCTACATCATTAAAGTAAAGTAAAG 778597
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 753 AACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 812
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778598 GACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 778657
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 813 TCAAAAAGCAAAATGATTCACACTTTGTTCAAAAGCAATTTGCGAGAGAGAGAGAGAGAG 872
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778658 GGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 778717
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 873 CATATCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 921
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778718 TATGTGAAGATTAAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 778766
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 14
ABN70712
ID ABN70712 standard; DNA; 894 BP.
XX
AC ABN70712;
XX
DT 01-JUL-2002 (first entry)
XX
DE Streptococcus polynucleotide seq ID NO 9337.
XX
KM Streptococcus; GMS; GBS; group B streptococcus; Streptococcus agalactiae;
XX group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
XX antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
OS Streptococcus agalactiae.
XX
XX WO200234771-A2.
XX
XX 02-MAY-2002.
XX
XX

```

```

PF 29-OCT-2001; 2001WO-GB04789.
XX
XX 27-OCT-2000; 2000GB-0026333.
PR 24-NOV-2000; 2000GB-0028727.
PR 07-MAR-2001; 2001GB-0005640.
XX
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
XX Telford J, Maignan V, Margalit Ros YI, Grandi G, Fraser C;
PI Tetelin H;
XX
XX WPI; 2002-352536/38.
XX
XX P-PDB; ABP30081.
XX
XX
XX New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein -
XX
XX Claim 7; Page 4058; 4525pp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
XX streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
XX the specification. The proteins have antibacterial and antiinflammatory
XX activity. (I), nucleic acids encoding (I), ABN65044-ABN71526 and
XX antibodies that bind (I) are used in the manufacture of medicaments for
XX the treatment or prevention of infection or disease caused by
XX Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX Nucleic acids encoding (I) are used to detect streptococcus in a
XX biological sample. (I) is used to determine whether a compound binds to
XX (I). A composition comprising (I) or a nucleic acid encoding (I), may be
XX used as a vaccine or diagnostic composition. The disease caused by
XX Streptococcus that is prevented or treated may be meningitis. Nucleic
XX acid encoding (I) may be used to recombinantly produce (I) and may be
XX used in gene therapy. Antibodies to (I) are used for affinity
XX chromatography, immunoassays, and distinguishing/identifying
XX Streptococcus proteins.
XX
XX
XX Sequence 894 BP; 356 A; 146 C; 155 G; 237 T; 0 other;
XX
XX
XX Query Match 26.4%; Score 255.4; DB 24; Length 894;
XX Best Local Similarity 57.7%; Pred. No. 2.6e-49;
XX Matches 501; Conservative 0; Mismatches 356; Indels 12; Gaps 2;
XX
XX
XX 56 TCACATATTATCAGTACGAACTTTAGCAGCTGTTC---GAAAGGTCAGAGAGTGCAG 112
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2 TGCACTTATGTCAGTTGGCCACATTTAGCTGCTCAGGTAAACATCAATGAGAACAA 61
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 113 ACCTTATCAGCATGAAGAGGAGTGTATTCAGAGACATCAATTTTATGAGCAAGTGAAGA 172
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 62 ATGTGTTTACAAATGAAGGGCGACACTTATACGTCTCTGATTTTATGATCAAGTAAAAA 121
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 173 GCAACCTTCAGCCCAACAAGCTTGTATTAATATGACATCCAAAAGTTTGTGAANAAC 232
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 122 CATCAAAAAGCTGCACAAACATCAATGCTTACATGATTCCTCTCAGCTGTTTGTATGAC 181
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 233 AATATGCTCAGAGCTTGTATGATTAAGAGAGTGTATGATTAATGCGAGAGAGAGAGAG 292
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 182 AGTATGATGATTAAGAGTTCAGATTAAGAGAGTTCAGAGAGTTCATTAAGAGAGTAAAG 241
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 293 AATATGCGCAAAACTACCAAGCTGTCTGTACACAGAGATATGACTTGTGAAGACGTA 352
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 242 GCTATGCTAATTCATTTTCAACGCGACTTTCACAAAGAGTTTGAATCCGGAAGGTTTACA 301
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 353 AAGCTCAAAATTCGTAACAGTAAATTTAGTGTGAGTGGCACTTAAGAGAGTACAGAGCTG 412
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 302 AACACAAATTCGACACACTATGCTTGTGATATCTCTTAAGAGAGAGAGAGAGAGAGAG 361
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 413 AATGACAGATGAAGCTTATGAAGAGAGCTTTGATGATGATGATGATGATGATGATGATGAT 472
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 362 AATTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 421
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

OY	473	AAATCATCGCTTAAATATGAAGATTAAGGCCAAAGAAAGTTCGCAAAAAGCCAAAGCGAG	532
Dd	422	AAGTAAATAAATTTGGATGCAGAGATTAAGCTAAATCTGTCTTAAAGATGTAAGGCTG	481
OY	533	AAGGTCTGATTTTGCTCAATTAGCCAAAGATTAATTCACACTGATGAAAAAACAAAGAAA	592
Dd	482	ATGGAGCTGATTTTGCAAAAGATTGCCAAAAGAAAAAACACAGCTACTGATTAAGAAAGTT	540
OY	593	ATGTGGGGAATTAACCTTTGATTCTGTCTCAACAGAACTACCTGAGCAAGTCAAAAAAG	652
Dd	541	-----GACTATATAAATTTTGATTCTGTGACGGGACAACCCCTCCCTAAAGAAATTGTCTAG	592
OY	653	CCGCTTTCGCTTTAGATGATGATGGTGTTCGTAGTGTGATTAACAGCACTGGCACAAG	712
Dd	593	CAGCCTTTAAGCTAGATATAAAATGGTGTTCAGATGTGTGTTCAACGGTGTGATTCACAA	652
OY	713	CCTACAGTAGCCAAATATTACATTGTTPAAACCTCACTAAGAAAACGAAAAATCTGATA	772
Dd	653	CTTATATAAACCAAGTACTACATCATATAAAGTACGATTAAGACAGAGAAAAATCTGATT	712
OY	773	TTGATGACTACAAAGAAAAATTAAAAACGTATATCTGTGACTCAAAAACAAAATGATTCAA	832
Dd	713	GGAATCTTACAAAAATGCGTTAAAGAAAGTATTCTTAAGGATTAACACAGCATAGAG	772
OY	833	CATTGTTCCAAAGCATTTATCGGAAAAAGATTTCGAAGCAGCCCAATATCAAGGTTAAGACC	892
Dd	773	CCTTCCAAATATAAGTGAATTTCAAAAAGCCTTGAAAAAGCTAATGTTAGATTAAGATA	832
OY	893	AAGCCTTCCAAAATATCTTTACCCCAATAT	921
Dd	833	AAAGCTTTGACAGGCATCTTATCACAATAT	861
RESULT 15			
AAAX20090			
ID	AAAX20090 standard. DNA: 1110 BP.		
AC	AAAX20090;		
XX	20-APR-1999 (first entry)		
DE	Enterococcus faecalis gene EF049.		
KM	Enterococcus faecalis; infection; vaccine; immune response; diagnosis;		
KW	detection; attenuation; antigenic; ss.		
OS	Enterococcus faecalis.		
XX	WO9850554-A2.		
PN	12-NOV-1998.		
XX	04-MAY-1998; 98WO-US08959.		
PE	14-NOV-1997; 97US-0066009.		
PR	06-MAY-1997; 97US-0044031.		
PR	16-MAY-1997; 97US-0046655.		
XX	(HDMA-) HUMAN GENOME SCI INC.		
XX	Bailey C, Choi GH, Hromockyj A, Kunsch CA;		
XX	WPI: 1999-070095/06.		
XX	P-PSDB; AAY00100.		
PT	New isolated Enterococcus faecalis polynucleotides - used to develop		
PT	products for the detection of Enterococcus and for use in vaccines		
PT	for prevention or attenuation of Enterococcus infection		
XX	Claim 1; Page 126; 301pp; English.		
XX	The present sequence represents a gene isolated from		
CC	Enterococcus faecalis. The present invention describes genes, proteins		

Query Match	22.4%	Score 216.2	DB 20	Length 1110
Best Local Similarity	55.4%	Pred. No. 3e-40		
Matches	513	Conservative	0	Mismatches 389; Indels 24; Gaps 4
CC	and antigenic polypeptides isolated from <i>E. faecalis</i> . The proteins can			
CC	be used in vaccines for preventing or attenuating an infection caused			
CC	by a member of the Enterococcus genus in an animal. They can also be			
CC	used for detecting Enterococcus antibodies in a sample. The nucleotide			
CC	sequences can be used for detecting Enterococcus nucleic acids.			
CC	Products from the present invention can also be used for screening			
CC	compounds to identify agonists and antagonists of <i>E. faecalis</i> protein			
CC	activity.			
XX	Sequence 1110 BP; 471 A; 191 C; 201 G; 244 T; 3 other;			
SO				
QY	ATGAAAGAAAAATTTATTGGCAGGTGCGATCACAATTTATCAGTAGCAACTTTAGACGT	87		
DB				
QY	82 ATGAAAGAAAAAACTAATCTTACTGCTCAGCGGCGCATGGCGTTTGTGTTAGCAGG	141		
DB				
QY	88 TGTTCGAAAGGGTGCAGAAAGTGCAGACCTTATCAGATGAAGGGGATGTCTATACAGA	147		
DB				
QY	142 TGTTCAGCGCTCAA-----AGATATCGCAACATGAAGTTCAACANTTCTGT	195		
DB				
QY	148 CATCAATTTTATGACAAAGTAAAGCAACCTTCAGCCCAACAGCTGTTTAAATATG	207		
DB				
QY	196 GATGATTTTATTAACCAATTTAAGAACCAACACATGACCAACACGCTTTAGCCAAATG	255		
DB				
QY	208 ACCATCCAAAAGTTTGTGAAACAAATATGGCTGACCTTGATATGAAGAGGTGAT	267		
DB				
QY	256 GTTATTTTAAAGCTCTTTGACAAAAATATGGCGACAAAGTAACTACCAAGANNTCAA	315		
DB				
QY	268 GATACTATTGCGGAAGAAAAAACA-----ATATGGCGAAACGTAACCAACGTGC	318		
DB				
QY	316 AAAAATTGAGCAAGCCAAAGAACAGTAGAAGCAAGCGGGAAGTCTCTGATGCA	375		
DB				
QY	319 TTGTCACAAAGCATATGACTCTTGAAACACGTAAAGCTCAATTCGTACAGTAATTA	378		
DB				
QY	376 TTTAAACAAAGCTGGTTTAACTGTAACAAATTCGACGAAAGCTTTAAACAAAGCAGCC	435		
DB				
QY	379 GTTGAAGTTGGCAGTTAAGAAAGTAGCAGAGAGCTGATTTACAGATAGAGCTTAAAGAA	438		
DB				
QY	436 TATGATGACAGGTCTAA-----AGCCCACTTAAATTAAGATGAGACTTAAACAA	489		
DB				
QY	438 GCGTTTGAATGAGTACACTCCAGATGTACAGCGCTCAATATCCTTAAATGAAGAT	498		
DB				
QY	490 GCTTGGGCAAGTTTCCATCCAGAAAGTAGAAGCAAAATTAATCCAGTTGCTTCAGAGAT	549		
DB				
QY	499 AAGGCCAAAGAGTTCTCGAAAAGGCAAGGCAAGAGGCTGATTTGGTCATTAAGCC	558		
DB				
QY	550 GATGCCA---AAGCTCTCAAGAAAGAAATCCTGACGCGGCGATTTCACAAAAATTTGCT	606		
DB				
QY	559 AAGATTAATTCAGATGATGAAAAAACAAAAAGAAATGTGGAGAAATTAACCTTTGATCT	618		
DB				
QY	607 AAAGAAAAATTCACAGATCTCTGCTCAAGAAAAAGATGGCGGTAAATTTAAATTTGATTC	666		
DB				
QY	619 GCTTCAACGAAGTCCCTGAGCAAGTCAAAAAAGCGGCTTGGCTTAGATGGAGTGT	678		
DB				
QY	667 CAAGCAGACACTGTTCTCTCCGAGTTTAAAGAGCTGCTTCAATTTAAAGATGGCGAA	726		
DB				
QY	679 GTTCTGATGATGATTAACGCACTGGCAGACAGCTTACAGTAGAGCCAAATTTACATTGA	738		
DB				
QY	727 GTGTGAGAACCAATTCCTGCAACAATATGCAAACTACCAACCAACCTACTATGTAGTG	786		
DB				
QY	738 AAATCTACTTAAGAAAAACAAAAATATCATATATTTGATGACTACAAAGAAAAATTTAAA	798		
DB				
QY	787 AAAATGACAAAAAACAACAAAAAGGCAATGACATGAACCTTATATAAAAAGATCAAG	846		
DB				
QY	799 ACTGTTATCTTACTCAAAAAAATGATTCACATTTGTTCCAAAGCATATATCGGAAA	858		
DB				
QY	847 AAAATTTGCTGAAGAAACAAATTTAGCCGATCAACACTTTGTTTCGAAAGATCATATGAC	906		
DB				
QY	859 GAATTCAGACGCCAATATCAAGGTTTAAAGCAACCTTCCAAAATATCTTTACCAA	918		
DB				

Db 907 GAATTAAAGCGGCCCAATGTGAAATTTAAAGATGATGCCCTTCAAGACGCTTTAGCAGGC 966
 Qy 919 TATATCGGTGTGTGAGATTCAAGCTC 944
 |||||
 Db 967 TACATGCAGAACTGAATCTTCAAGCGC 992

Search completed: September 4, 2003, 19:26:26
 Job time : 325 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2003, 19:20:57 : Search time 87 Seconds
(without alignments)
4900.875 Million cell updates/sec

Title: US-10-049-473A-1

Perfect score: 966

Sequence: 1 agtaacactatctcaaga.....gcagtagtacatcaacga 966

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq: *
6: /cgn2_6/ptodata/1/ina/Backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	966	100.0	11309	4	US-08-961-527-108 Sequence 108, App
2	552.4	57.2	640	3	US-08-961-083-33 Sequence 33, Appl
3	552.4	57.2	640	3	US-09-536-784-33 Sequence 33, Appl
4	265.6	27.5	791	4	US-08-858-207A-219 Sequence 219, App
5	216.2	22.4	1110	4	US-09-071-035-181 Sequence 181, App
6	209.8	21.7	1035	4	US-09-107-532A-2701 Sequence 2701, Ap
7	206.4	21.4	912	4	US-09-107-532A-2459 Sequence 2459, Ap
8	203.2	21.0	970	4	US-09-071-035-183 Sequence 183, App
9	151.2	15.7	909	4	US-09-107-532A-1082 Sequence 1082, Ap
c 10	59.8	6.2	7218	1	US-08-232-463-14 Sequence 14, Appl
11	52	5.4	3095	6	5231168-1 Patent No. 5231168
12	48.8	5.1	4673	5	US-07-638-431-1 Sequence 1, Appl
13	48.8	5.1	4673	5	PCT-US92-00018-1 Sequence 578, App
14	46	4.8	633	4	US-09-134-001C-578 Sequence 1407, Ap
15	45.2	4.7	1239	4	US-09-134-001C-1407 Sequence 3, Appl
16	44.4	4.6	1891	3	US-08-973-462-3 Sequence 600, App
17	44.4	4.5	1155	4	US-09-134-001C-600 Sequence 67, Appl
18	43.4	4.5	7163	4	US-08-961-527-67 Sequence 93, Appl
19	43	4.5	835	3	US-08-961-083-93 Sequence 93, Appl
20	43	4.5	835	3	US-09-536-784-93 Sequence 93, Appl
21	43	4.5	1714	3	US-08-961-083-3 Sequence 3, Appl
22	43	4.5	1714	3	US-09-536-784-3 Sequence 3, Appl
23	43	4.5	6201	2	US-08-790-912-1 Sequence 1, Appl
c 24	43	4.5	15213	4	US-08-961-527-26 Sequence 26, Appl
25	42	4.3	1269	4	US-09-601-198-178 Sequence 178, App
26	41.6	4.3	5361	3	US-08-973-462-2 Sequence 2, Appl
27	41.6	4.3	6152	3	US-08-973-462-1 Sequence 1, Appl

28	41.4	4.3	993	4	US-09-134-001C-974 Sequence 974, App
c 29	41	4.2	10993	2	US-08-961-527-15 Sequence 15, Appl
30	40.8	4.2	3666	4	US-08-682-517-13 Sequence 13, Appl
31	40.8	4.2	3666	2	US-08-682-517-14 Sequence 14, Appl
32	40.8	4.2	4197	2	US-08-682-517-7 Sequence 7, Appl
33	40.8	4.2	4197	2	US-08-682-517-8 Sequence 8, Appl
34	40.8	4.2	4766	5	PCT-US93-07261-10 Sequence 10, Appl
c 35	40.6	4.2	729	4	US-09-601-198-16 Sequence 16, Appl
36	40.4	4.2	921	2	US-08-795-475-2 Sequence 2, Appl
37	40.4	4.2	1305	4	US-09-328-234A-6 Sequence 6, Appl
38	40.4	4.2	1308	2	US-08-795-475-4 Sequence 4, Appl
39	40.4	4.2	3222	4	US-08-714-741-39 Sequence 39, Appl
40	40.4	4.2	3227	1	US-08-676-967-2 Sequence 2, Appl
41	40	4.1	2277	1	US-08-676-974-2 Sequence 2, Appl
42	40	4.1	2277	2	US-09-098-487-2 Sequence 2, Appl
c 43	40	4.1	3534	4	US-09-134-001C-2269 Sequence 2269, Ap
44	39.8	4.1	1374	4	US-09-601-198-131 Sequence 131, App
45	39.6	4.1	480	4	US-09-134-001C-914 Sequence 914, App

ALIGNMENTS

RESULT 1

US-08-961-527-108
; Sequence 108, Application US/08961527

; Patent No. 6420135

; GENERAL INFORMATION:

APPLICANT: Charles Kunach

TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

NUMBER OF SEQUENCES: 391

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage

OPERATING SYSTEM: HP Vectra 486/33

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,527

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 108:

SEQUENCE CHARACTERISTICS:

LENGTH: 11309 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-961-527-108

Query Match 100.0%; Score 966; DB 4; Length 11309;

Best Local Similarity 100.0%; Pred. No. 7.4e-232;

Matches 966; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTAACCTTCTTCCTCAAGAGAGTACATGAGAAATATGCGCATGCA 60

DB 7632 AGTAACCTTCTTCCTCAAGAGAGTACATGAGAAATATGCGCATGCA 7691

QY	61	CGATTATAGTGCACACTTTAGCACTTGTTCGAAAGGTCAGAAAGTGCAGACTTATC	120
Db	7692	CTATTATCAGTAGCAACTTTAGCACCTTGTTCGAAAGGTCAGAAAGTGCAGACTTATC	7751
QY	121	AGCATGAAAAGGGGATGCTATTACAGAACATCAATTTATGAGCAAGTAAAAAGCAACCT	180
Db	7752	AGCATGAAAAGGGGATGCTATTACAGAACATCAATTTATGAGCAAGTAAAAAGCAACCT	7811
QY	181	TCAGGCCCAACAGTCTTGTGTTAAATATGACCATCCAAAAGTTTTTGA AAAACAATATGCG	240
Db	7812	TCAGGCCCAACAGTCTTGTGTTAAATATGACCATCCAAAAGTTTTTGA AAAACAATATGCG	7871
QY	241	TCAGAGCTTGATGATNAAGAGGTTGATATATCTATTGGCGAAGAAAAACAATATGCG	300
Db	7872	TCAGAGCTTGATGATNAAGAGGTTGATATATCTATTGGCGAAGAAAAACAATATGCG	7931
QY	301	GAAACTACCAACGCTGTTGTCACACAGAGTATGACTCTTGAACACGTAAAGCTCAA	360
Db	7932	GAAACTACCAACGCTGTTGTCACACAGAGTATGACTCTTGAACACGTAAAGCTCAA	7991
QY	361	ATTCGTACAGTAATTTAGTTGAGTTGGCAGTTAAGAGTNGCAGAGCTGAATGACA	420
Db	7992	ATTCGTACAGTAATTTAGTTGAGTTGGCAGTTAAGAGTNGCAGAGCTGAATGACA	8051
QY	421	GATGAGCGCTTATAGAAAAGCTTTGATGAGTACACTCCAGATGTAAAGCGCTCAATCATC	480
Db	8052	GATGAGCGCTTATAGAAAAGCTTTGATGAGTACACTCCAGATGTAAAGCGCTCAATCATC	8111
QY	481	CGTCTTAATATGAAGATPAAGGCCAAAGATCTCCGAAAAAGCCAGAGGAGTGCT	540
Db	8112	CGTCTTAATATGAAGATPAAGGCCAAAGATCTCCGAAAAAGCCAGAGGAGTGCT	8171
QY	541	GATTTTGTCTCAATTAGCCAAAAGATPATTCACCTGATGAAAAAACAAAAAGAAATGTGGA	600
Db	8172	GATTTTGTCTCAATTAGCCAAAAGATPATTCACCTGATGAAAAAACAAAAAGAAATGTGGA	8231
QY	601	GAAATTAACCTTGATTCGTTCAACAGAACTACTGTGCAAGTCAAAAAACCGCGCTTC	660
Db	8232	GAAATTAACCTTGATTCGTTCAACAGAACTACTGTGCAAGTCAAAAAACCGCGCTTC	8291
QY	661	GCTTTAGATGTGATGTGTGTTTCGATGTGATTCACAGCACTGGCAGACAAAGCTTACGT	720
Db	8292	GCTTTAGATGTGATGTGTGTTTCGATGTGATTCACAGCACTGGCAGACAAAGCTTACGT	8351
QY	721	AGCCATATTTCAATGTGTAACCTACATAAGAAAAACA AAAAATATCTAATATTGATGAC	780
Db	8352	AGCCATATTTCAATGTGTAACCTACATAAGAAAAACA AAAAATATCTAATATTGATGAC	8411
QY	781	TACAAAGAAAAATTA AAAACTGTATCTTGACTCAAAAACAAAATGATTCAACTTTGTT	840
Db	8412	TACAAAGAAAAATTA AAAACTGTATCTTGACTCAAAAACAAAATGATTCAACTTTGTT	8471
QY	841	CAAGCAATTAATCGGAAAAAGAAATTCAGACGCCAATATCAAGGTTAAAGCAAGCCCTTC	900
Db	8472	CAAGCAATTAATCGGAAAAAGAAATTCAGACGCCAATATCAAGGTTAAAGCAAGCCCTTC	8531
QY	901	CAAAATATCTTTACCACATATATGCGTGTGAGATTCAGAGCTCAACAGTACAGTACATCA	960
Db	8532	CAAAATATCTTTACCACATATATGCGTGTGAGATTCAGAGCTCAACAGTACAGTACATCA	8591
QY	961	AACGAA 966	
Db	8592	AACGAA 8597	

RESULT 2
 US-08-961-083-33
 ; Sequence 33, Application US/08961083
 ; Patent No. 6159469
 ; GENERAL INFORMATION:
 ; APPLICANT: Chol et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
 ; NUMBER OF SEQUENCES: 452

	CORRESPONDENCE ADDRESS:	
	ADDRESSEE: Human Genome Sciences, Inc.	
	STREET: 9410 Key West Avenue	
	CITY: Rockville	
	STATE: Maryland	
	COUNTRY: USA	
	ZIP: 20850	
	COMPUTER READABLE FORM:	
	MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage	
	COMPUTER: HP Vectra 486/33	
	OPERATING SYSTEM: MSDOS version 6.2	
	SOFTWARE: ASCII Text	
	CURRENT APPLICATION DATA:	
	APPLICATION NUMBER: US/08/961,083	
	FILING DATE:	
	CLASSIFICATION: .435	
	PRIOR APPLICATION DATA:	
	APPLICATION NUMBER:	
	FILING DATE:	
	ATTORNEY/AGENT INFORMATION:	
	NAME: Brookes, A. Anders	
	REGISTRATION NUMBER: 36,373	
	REFERENCE/DOCKET NUMBER: PB340P2	
	TELECOMMUNICATION INFORMATION:	
	TELEPHONE: (301) 309-8504	
	TELEFAX: (301) 309-8512	
	INFORMATION FOR SEQ ID NO: 33:	
	SEQUENCE CHARACTERISTICS:	
	LENGTH: 640 base pairs	
	TYPE: nucleic acid	
	STRANDEDNESS: double	
	TOPOLOGY: linear	
	US-08-961-083-33	
Query Match	57.2% Score 552.4; DB 3 Length 640;	
Best Local Similarity	94.7%; Pred. No. 5.4e-129;	
Matches 595; Conservative	0; Mismatches 26; Indels 7; Gaps 2	
OY	90 TTCGAAAGGTCAGAAGGTGCAGACCTTATCAGCATGAAGGGGATGTCATTACGAAACA 149	
Db	1 TTGCAAAAGGTCAGAAAGGTGCAGACCTTATCAGCATGAAGGGGATGTCATTACGAAACA 60	
OY	150 TCAATTTATAGCAAGTAAGAAAGCAACCCCTTCAGCCCAACAAGCTTGTTAATAATGAC 209	
Db	61 TCAATTTATAGCAAGTAAGAAAGCAACCCCTTCAGCCCAACAAGCTTGTTAATAATGAC 120	
OY	210 CATCCAAAAGTTTTTGAAAAAACAATATAGCGTCAGAGCTTGATGTAAGAAGGTGATGA 269	
Db	121 CATCCAAAAGTTTTTGAAAAAACAATATAGCGTCAGAGCTTGATGTAAGAAGGTGATGA 180	
OY	270 TACTATTGCCAGAAAAAACATATATGCGAAAACTACCAACGCTGCTTGTCAACAAGC 329	
Db	181 TACTATTGCCAGAAAAAACATATATGCGAAAACTACCAACGCTGCTTGTCAACAAGC 240	
OY	330 AGGATATGACTTTGAAACAGCTAAAGCTCAAATTCGTCAAGTAAATAGTTAGTTAGTTGGC 389	
Db	241 AGGATATGACTTTGAAACAGCTAAAGCTCAAATTCGTCAAGTAAATAGTTAGTTAGTTGGC 300	
OY	390 AGTTAAGAAGGTAGCAGAAAGCTGAATTGACAGATGAAGCTCTATAAGAAAGCCTTGATGA 449	
Db	301 AGTTAAGAAGGTAGCAGAAAGCTGAATTGACAGATGAAGCTCTATAAGAAAGCCTTGATGA 360	
OY	450 GTAACATTCAGATGTAACGGCTCAAATTCATCCGCTTAAATATGAGATAAGCCCAAGA 509	
Db	361 GTAACATTCAGATGTAACGGCTCAAATTCATCCGCTTAAATATGAGATAAGCCCAAGA 420	
OY	510 AGTCTCTGAAAAAACCAGAGAGAGGCTGATTTTGCCTAATTAAGCCAAGATPAATTC 569	
Db	421 AGTCTCTGAAAAAACCAGAGAGAGGCTGATTTTGCCTAATTAAGCCAAGATPAATTC 480	
OY	570 AACTGATGAAAAAACAAGAAATGTTGAGAAATTAACCTTGATTCCTGCTCAACAGA 629	
Db	481 AACGATGAAAAAACAAGAAATGTTGAGAAATTAACCTTGATTCCTGCTCAACAGA 540	

Oy 630 AGCAATCGT---AGCAATCTAAAAACCCGCTTTTCGCTTAGATGTGGATGTTTGAT 687
 ||||| |
Db 541 AGTACTGGAGCAAGTCCAAAACCCGCTTTTCGCTTTAGATGTGGATGTGTTCT 600
 ||||| |
Oy 688 G-----TGATTACGCACAATCGCACACA 710
 | ||||| |
Db 601 GGATGTGATTACAGCAACTGGGGCACA 628
 | ||||| |

RESULT_3
US-09-536784-33
Sequence 33, Application US/09536784
Patent No. 6573082

RESULT 3

```

: Sequence 33, Application US/09536784
: Patent No. 6573082
:
: GENERAL INFORMATION:
: APPLICANT: Choi et. al.
: TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
: NUMBER OF SEQUENCES: 452
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Human Genome Sciences, Inc.
: STREET: 9410 Key West Avenue
: CITY: Rockville
: STATE: Maryland
: COUNTRY: USA
: ZIP: 20850
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
: COMPUTER: HP Vectra 486/33
: OPERATING SYSTEM: MSDOS version 6.2
: SOFTWARE: ASCII Text
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/536,784
: FILING DATE: 30-Oct-1997
: CLASSIFICATION: <Unknown>
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/961,083
: FILING DATE: OCT-30-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Michelle S. Marks
: REGISTRATION NUMBER: 41,971
: REFERENCE/DOCKET NUMBER: PB340P3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 309-8504
: TELEFAX: (301) 309-8512
:
: INFORMATION FOR SEQ ID NO: 33:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 640 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
:
: SEQUENCE DESCRIPTION: SEQ ID NO: 33:
:
: US-09-536-784-33

```

Query Match	57.28	Score	552.4	DB 4	Length	640			
Best Local Similarity	94.78	Pred	No. 5.4e-129						
Matches	595	Conservative	0	Mismatches	26	Indels	7	Gaps	2

QY	90	TTGCAAAAGGCTCGAAGGTGTCAGACCTTATCAGCATGAAAGGGATGTCATTACAGACA	149
Db	1	TTGCAAAAGGCTCGAAGGTGTCAGACCTTATCAGCATGAAAGGGATGTCATTACAGACA	60
QY	150	TCAATTTTATGACCAAGTAAAAAGCAACCCCTTCAGCCCAACAGTCTGTTAAATATGAC	209
Db	61	TCAATTTTATGACCAAGTAAAAAGCAACCCCTTCAGCCCAACAGTCTGTTAAATATGAC	120
QY	210	CATCCAAAAGTTTGGAAAAACATATGGCTCAGAGCTTGATGATTAAGAGGTTGATGA	265
Db	121	CATCCAAAAGTTTGGAAAAACATATGGCTCAGAGCTTGATGATTAAGAGGTTGATGA	180
QY	270	TACATTTGCCGAGAAAAAAACAATATGGCGAAATACCAAGCTGCTTGACACAGC	329
Db	181	TACATTTGCCGAGAAAAAAACAATATGGCGAAATACCAAGCTGCTTGACACAGC	240

Oy	330	AGGAATGACCTTGGAAACAGTAAAGCCTAAATTCGTCAAGTAAATTAGTTAGTTGGC	389
Dp	241	AGGTATGACTCTGGAAACAGTAAAGCCTAAATTCGTCAAGTAAATTAGTTAGTTGGC	3000
Oy	390	AGTTAAGAAAGTAGCAGAAAGCTGAATTGACAGATGAACCTATATAAGAAAGCCTTTGATGA	449
Dp	301	AGTTAAGAAAGTAGCAGAAAGCTGAATTGACAGATGAACCTATATAAGAAAGCCTTTGATGA	3600
Oy	450	GTACACTCCAGATGTAAAGGCTCAATCATCCGCTTTAATATGAAGTAAAGGCCAAGA	508
Dp	361	GTACACTCCAGATGTAAAGGCTCAATCATCCGCTTTAATATGAAGTAAAGGCCAAGA	4200
Oy	510	AGATTCTCGAAAAGGCCAAGCAGAAAGGTCTGATTTGCTCAATTAGCCAAAGATTAATTC	569
Dp	421	AGTTCTCGAAAAGGCCAAGCAGAAAGGTCTGATTTGCTCAATTAGCCAAAGATTAATTC	4800
Oy	570	AACGTATGAAAAACAAAAAGAAATGTGTGAGAAATTAACCTTGTATCTCGCTTCACAGA	629
Dp	481	AACGTATGAAAAACAAAAAGAAATGTGTGAGAAATTAACCTTGTATCTCGCTTCACAGA	5400
Oy	630	AGTACCTG--AGCAAGTCAAAAAAGCCGCTTCGCTTAGATGTGTGATGTGTTCGAT	687
Dp	541	AGTACCTGAGCAAGTCAAAAAAGCCGCTTCGCTTAGATGTGTGAGTGTGTTC	6000
Oy	688	G-----TGATTACAGCACTGGAGCA	710
Dp	601	GGATGTGATTACAGCACTGGAGCA	628

RESULT 4

US-08-858-207A-219
 ? Sequence 219, Application US/08858207A
 ? Patent No. 6348328
 ? GENERAL INFORMATION:
 ? APPLICANT: Black, Michael
 ? APPLICANT: Hodgson, John
 ? APPLICANT: Knowles, David
 ? APPLICANT: Nicholas, Richard
 ? APPLICANT: Stodola, Robert
 ? TITLE OF INVENTION: No. 6348328el Compounds
 ? NUMBER OF SEQUENCES: 552
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSEE: Smithkline Beecham Corporation
 ? STREET: 709 Swedeland Road
 ? CITY: King of Prussia
 ? STATE: PA
 ? COUNTRY: USA
 ? ZIP: 19406-0939
 ? COMPUTER READABLE FORM:
 ? MEDIUM TYPE: Diskette
 ? COMPUTER: IBM Compatible
 ? OPERATING SYSTEM: DOS
 ? SOFTWARE: FASTSEQ for Windows Version 2.0
 ? CURRENT APPLICATION DATA:
 ? APPLICATION NUMBER: US/08/858.207A
 ? FILING DATE: 09-May-1997
 ? CLASSIFICATION: 435
 ? PRIOR APPLICATION DATA:
 ? APPLICATION NUMBER: 60/017670
 ? FILING DATE: 14-May-1996
 ? ATTORNEY/AGENT INFORMATION:
 ? NAME: Gilm1, Edward R
 ? REGISTRATION NUMBER: 38,891
 ? REFERENCE/DOCKET NUMBER: P50475
 ? TELECOMMUNICATION INFORMATION:
 ? TELEPHONE: 610-270-4478
 ? TELEFAX: 610-270-5090
 ? TELEX:
 ? INFORMATION FOR SEQ. ID NO.: 219:
 ? SEQUENCE CHARACTERISTICS:
 ? LENGTH: 791 base pairs
 ? TYPE: nucleic acid
 ? STRANDEDNESS: single

TOPOLOGY: linear
US-08-858-207A-219

Query Match 27.5%; Score 265.6; DB 4; Length 791;
Best Local Similarity 97.9%; Pred. No. 2.6e-57;
Matches 279; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 1 AGTAACTATTATCTCAAGAGTAGACATGAGAAAAAATATATGGCAGGTGCCATCACA 60
DB 507 AGTAACTATTATCTCAAGAGTAGACATGAGAAAAAATATATGGCAGGTGCCATCACA 566
QY 61 CTATTATGATGACACTTAGACGTTGTGCAAGGCGCAAGAGTGCAGACCTTATC 120
DB 567 CTATTATGATGACACTTAGACGTTGTGCAAGGCGCAAGAGTGCAGACCTTATC 626
QY 121 AGCATGAAAGGGAGTGTCAATGACAAATCAATTTATGAGCAATGAGAAAGCAACCT 180
DB 627 AGCATGAAAGGGAGTGTCAATGACAAATCAATTTATGAGCAATGAGAAAGCAACCT 686
QY 181 TCAGCCCAACAGTCTGTTAAATATGACCATCCAAAAGTTTGG-AAAAACAATATG 239
DB 687 TCAGCCCAACAGTCTGTTAAATATGACCATCCAAAAGTTTGGAAAAACAATATG 746
QY 240 CTCAGAGCTTGATGATAAGAGGTGATGATCTATTGCGCGAGA 284
DB 747 CTCAGAGCTTGATGATAAGAGGTGATGATCTATTGCGCGAGA 791

RESULT 5

US-09-071-035-181

Sequence 181, Application US/09071035

Patent No. 6448043

GENERAL INFORMATION:

APPLICANT: Gail H. Choi

TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

NUMBER OF SEQUENCES: 496

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/071,035

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: A. Anders Brookes

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB369P2

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 181:

SEQUENCE CHARACTERISTICS:

LENGTH: 1110 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-09-071-035-181

Query Match 22.4%; Score 216.2; DB 4; Length 1110;
Best Local Similarity 55.4%; Pred. No. 6.3e-45;
Matches 513; Conservative 0; Mismatches 389; Indels 24; Gaps 4;

QY 28 ATGAGAAAAAATTTATGGCAGGTGCCATCACACTATATAGTAGCACTTTAGACGCT 87
DB 82 ATGAGAAAAAATTTATGGCAGGTGCCATCACACTATATAGTAGCACTTTAGACGCT 141
QY 88 TGTTCGAAAGGCTCAGAGGTGCAGACCTTATCAGCATGAAAGGGAGTGCATTACGAGA 147
DB 142 TGTTCGAAAGGCTCAGAGGTGCAGACCTTATCAGCATGAAAGGGAGTGCATTACGAGA 195
QY 148 CATCATTTTATGAGCAAGTAGAAGCAACCTTACAGCCCAACAGTCTTTGAATATG 207
DB 196 GATGATTTTATGAGCAAGTAGAAGCAACCTTACAGCCCAACAGTCTTTGAATATG 255
QY 208 ACCATCCAAAAGTTTGTGAAAAACAAATATGCTCAGACCTTATGATGATGATGATGAT 267
DB 256 GTTATTTATTAAGTCTTTTAAAGAAAAATATGCGCAAGTAAGTACACAAATTTCA 315
QY 268 GATPACTATTCGCGAAGAAAAAACA-----ATATGGGAAAAACTACCAAGCTGTC 318
DB 316 AAAAATTTGACGAGGAGCCAAAGACAGTGAAGACAAAGGCGAAGTCTCTGATGCA 375
QY 319 TTGTCAAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 378
DB 376 TTTAAACAAAGCTGTTTAACTGAAAAACATTTCAAGAAACAGTTTAAACAAAGCAGCC 435
QY 379 GTTGAAGTGGCAGTTAAGAGGTGACGAGAGCTGATGATGATGATGATGATGATGAT 438
DB 436 TATGATGACAGTCTTAA-----AGCCCACTTAAATATGATGATGATGATGATGAT 489
QY 439 GCCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 498
DB 490 GCTTGGGCAAGTTTCCATTCACAGAGTGAAGACAAATTTATCAAGTGTCTTCAAGAGT 549
QY 499 AAGCCCAAGAGTCTGAAAAAGCCAGAGGAGGAGTGTGATTTGCTCAATATAGCC 558
DB 550 GATGCCA--AAGCTGCAAGAAAGAAATCACTGCGGGGAGGATTCACAAAATGCT 606
QY 559 AAGATTAATTCACATGATGAAAAAACAAGAAATGTTGGAGAAATTAATTAATTAAT 618
DB 607 AAGAAAAATTCACATGATGAAAAAACAAGAAATGTTGGAGAAATTAATTAATTAAT 666
QY 619 GCTTCACAGAGAGTCTGAGCAAGTCAAAAAAGCCGCTTTCGCTTATGATGATGATGAT 678
DB 667 CAAGCAACAGAGTCTTCCGCGAATTTAAGAGAGTGTCTTCAATTTAAGAGATGCGAA 726
QY 679 GTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 738
DB 727 GTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 786
QY 739 AAGCTCAGTAAAGAAAGAAAGAAATATGATGATGATGATGATGATGATGATGATGAT 798
DB 787 AAGATGAGAGAAAGAAAGAAAGAAATATGATGATGATGATGATGATGATGATGATGAT 846
QY 799 ACTGTTATCTGACTCAAAAAACAATATGATGATGATGATGATGATGATGATGATGAT 858
DB 847 AAAATTTGCTGAAGAAACAAATTTAGCGATCAAAATTTGTTGCAAGATCATTAAGTAC 906
QY 859 GAATTCGAGAGCAAGCAATATCAAGTTAAGACCAAGCTTCCAAAATATCTTACCAA 918
DB 907 GAATTTAAAGGCGCCAAAGTGAATTAAGATGATGATGATGATGATGATGATGATGAT 966
QY 919 TATATCGGTGGTGAAGATTCAGATC 944
DB 967 TACATGCAAACTGATTCACAGGC 992

RESULT 6

US-09-107-532A-2701

Sequence 2701, Application US/09107532A

Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

```

1      NUMBER OF SEQUENCES: 7310
2      ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
3      CORRESPONDENCE ADDRESS:
4      ADDRESSEE: GENOME THERAPEUTICS CORPORATION
5      STREET: 100 Beaver Street
6      CITY: Waltham
7      STATE: Massachusetts
8      COUNTRY: USA
9      ZIP: 02354
10
11     COMPUTER READABLE FORM:
12     MEDIUM TYPE: CD-ROM ISO9660
13     COMPUTER: PC
14     OPERATING SYSTEM: <Unknown>
15     SOFTWARE: ASCII
16
17     CURRENT APPLICATION DATA:
18     APPLICATION NUMBER: US/09/107,532A
19     FILING DATE: 30-Jun-1998
20
21     PRIOR APPLICATION DATA:
22     APPLICATION NUMBER: 60/085,598
23     FILING DATE: 14 May 1998
24     APPLICATION NUMBER: 60/051571
25     FILING DATE: July 2, 1997
26
27     ATTORNEY/AGENT INFORMATION:
28     NAME: Arinello, Pamela Deneke
29     REGISTRATION NUMBER: 40,489
30     REFERENCE/DOCKET NUMBER: GTC-012
31     TELECOMMUNICATION INFORMATION:
32     TELEPHONE: (781)893-5007
33     TELEFAX: (781)893-8277
34
35     INFORMATION FOR SEQ ID NO: 2701:
36
37     SEQUENCE CHARACTERISTICS:
38     LENGTH: 1035 base pairs
39     TYPE: nucleic acid
40     STRANDEDNESS: double
41     TOPOLOGY: circular
42
43     MOLECULE TYPE: DNA (genomic)
44
45     HYPOTHEetical: NO
46
47     ANTI-SENSE: NO
48
49     ORGANISM: Enterococcus faecium
50
51     FEATURE:
52     NAME/KEY: misc feature
53     LOCATION: (B) LOCATION 1..1035
54     SEQUENCE DESCRIPTION: SEQ ID NO: 2701:
55
56     US-09-107-532A-2701
57
58     Query Match          21.7%; Score 209.8; DB 4; Length 1035;
59     Best Local Similarity 54.8%; Pred. No.2,4e-43;
60     Matches 484; Conservative 0; Mismatches 387; Indels 12; Gaps 3;
61
62     QY      31  AAGAAAAAATTATTGGCAGGTGCCATCATCACTATTATCAGTAGCAACTTTAGCAGCTTGT 90
63           || ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
64     Db      31  AAAAAAACCAATTATTTTGGCAGCTACAGTGCCTGGCTGTTTAACTTACGACGCTTGT 90
65
66     QY      91  TCGAAGAGGTCAGAGGTCGAGACTCTTATCGCATGGAAGGGGAGATGTCTATACAGAACAT 150
67           || ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
68     Db      91  TC---AGCGCACTACTAATTAAGATATGCGAACGATGGAAGCGGAGACGATCCTGTATCT 147
69           || ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
70
71     QY      151 CAATTATTATGAGCAAGTGAAGAAAGCAACCCTTCAGCCCAACAACATGCTTGTTAATATGACC 210
72           || ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
73     Db      148 GATTTCATATGTATGAAGCAAAACTTGAATCATCGAACCAATCCTTAGTACAACGGATGATC 207
74
75     QY      211 ATCCAAAAGATTTTGAAGAACATATATGCGTCACAGCTGTGATGATAAGAGGCTGATGAT 270
76           || ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
77     Db      208 ATTATTAAGATTAATCAATACCAATATGCGCATTAAGTAACCGATTAAGCAAGTGGATGCA 267
78
79     QY      271 ACTATTCGCCGAAGAAAAAACCAATATATGCGCAAAACTACCAACGCTGTCTTGTCACAAGCA 330
80           || ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
81     Db      268 GAATACGCAACAACGAAGCAATCATTTAGGTGATCACTTTTGAATCAACAATTAAGAGCTGCT 327
82
83     QY      331 GGTATGACTCTTGAACACGTAAGGTCACAAATTTGTACAGTAAATTAAGTTGAAGTGGCA 390
84           || ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
85     Db      328 GGAATATCGAAGGATCTTATTAAGATATATTTTGTAAACAAT-----TTAGCTTTTGA 381

```

QY 391 GTTAAGAAAGTACAGAAAGCGTAATTCACAGATGAACCCATATAGAAGACCTTTGATGAG 450
 Db 382 GCTGGCTTGAAGGTCACGTAGATATTACAGATGACGATTTGAAAACCTGCTTGGAAATTC 441
 QY 451 TACACTCCAGATGTAAACGGCTCAATATCCGCTTATTATATAGATTAAGGCCAAGAA 510
 Db 442 TTCCATCCAGAGATGAGAGACCAAAATCATCAATATTACAGAGTGAAGATGAAGCAAAAAG- 500
 QY 511 GTTCTCGAAAAAGCCAAAGCAGAGAGTGCTGATTTTGCTCAATTAGCCAAAGATAATCA 570
 Db 501 --TGTGAAAAAATCAGCGGATGATGAGATGATTTCTCTAAATTAGCCAAAGACAATCA 558
 QY 571 ACTGATGAAAAAAGAAAAAGAAATGGTGGAGAAATTTACCTTATTTGCTTTCGTTCAACGAA 630
 Db 559 ACTGATTCACAGAGCAAAAAAGAGTGGGGGTAAAGTTAAATTCGATCAACGACAACT 618
 QY 631 GTACCTCAGCAAGTCAAAAAAGCCGCTTGCTTTAGATGTGATGGTGTGTTCTGATGTG 690
 Db 619 ATTCTCTCAGAGATGAAAAAGACGCTTTCAAAATGAAACGCTGAGATCTCGATGTT 678
 QY 691 ATTACAGCAACTGGCAGACAAAGCCTACAGTAGCCAAATATTACATTGTAAAACTCACTAAG 750
 Db 679 ATTACTCAACAAATCTCTACTTCTTATGCGACAGAAATACATCGTTGCAAAAAGTGTA 738
 QY 751 AAACACAAAAAATCATTAATTTGATGACTACAAAGAAAAAATTAATAAAGCTGATCTTG 810
 Db 739 AACCAAAATTAAGGCAATGACATGGATTAATTAACAGATCACTAAAGATATGCTACT 798
 QY 811 ACTCAAAAACAAAATGATTCACATTTTGTTCAAAAGCATTTACGSAAGAAAGATTGCAAGCA 870
 Db 799 GAACCTAAACGTAAGTGAACATGATTCACAACTTAAGTAATCGGTGAAGAAATTAAGAT 858
 QY 871 GCCAATATCAAGTTAAGGACCAAGCCTTGCAAAATATCTTTA 913
 Db 859 GCAAAATGTAAAAATCAAAAGATGATGACCTTGGAAAAATGTGTTGA 901

RESULT 7
 US-09-107-532A-2459
 : Sequence 2459, Application US/09107532A
 : Patent No. 6583275
 : GENERAL INFORMATION:
 : APPLICANT: Lynn A Doucette-Stamm and David Bush
 : TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 : ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
 : NUMBER OF SEQUENCES: 7310
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: GENOME THERAPEUTICS CORPORATION
 : STREET: 100 Beaver Street
 : City: Waltham
 : STATE: Massachusetts
 : COUNTRY: USA
 : ZIP: 02354
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: CD-ROM ISO9660
 : COMPUTER: PC
 : OPERATING SYSTEM: <Unknown>
 : SOFTWARE: ASCII
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/107,532A
 : FILING DATE: 30-Jun-1998
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 60/085,598
 : FILING DATE: 14 May 1998
 : APPLICATION NUMBER: 60/051571
 : FILING DATE: July 2, 1997
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Arinello, Pamela Deneke
 : REGISTRATION NUMBER: 40,489
 : REFERENCE/DOCKET NUMBER: GTC-012
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277
 INFORMATION FOR SEQ ID NO: 2459:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 912 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: circular
 MOLECULE TYPE: DNA (genomic)
 HYPOTHEICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Enterococcus faecium
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: (b) LOCATION 1...912
 SEQUENCE DESCRIPTION: SEQ ID NO: 2459:
 US-09-107-532A-2459

Query Match 21.4%; Score 206.4; DB 4; Length 912;
 Best Local Similarity 55.3%; Pred. No. 1.7e-42;
 Matches 490; Conservative 0; Mismatches 381; Indels 15; Gaps 4;

QY 28 ATGAGAAAAAATTTATGGCAGGTGCGCATCACATATTATCAGTACCACTTTAGCAGCT 87
 DB 1 ATGAAAAAAGATTTTATGACATCTTGCCATCTAGCGACCTGGCTTA--TTGTACAGC 57
 QY 88 TGTTCGAAAGGCGCAGAGGTGACAGCTTATCAGCATGAAAGGAGTGTCTTACAGAA 147
 DB 58 TGTACGAATGCCGCTGAAAAACAGCGCT--AGTTATTAAGCGGTACATCAACAGCGAA 114
 QY 148 CATCAATTTTATGAGCAAGTAAAGCAACCCCTTCAGCCCAACAGTCTTGTAAATATAG 207
 DB 115 CAAAGATATGATGCTTGAAGAAAAATGCAAGAGCAGATTCGCTGTGCAACGTTG 174
 QY 208 ACCATCCAAAAATTTTGAATAATATGCTCAGAGCTTGATTAAGAGCTTGTAT 267
 DB 175 ATCGTTATCAAGTATTGTAAGTAATATGAGATGACGTATCAACCAAGAAATCGAT 234
 QY 268 GATCTATTGGCCGAGAAAAAACAATATGGGAAAGTACCAAGCTGTCTTGACAA 327
 DB 235 TCTCATATGATCAACCAAGAAACATTTGGGATTCATTGATTCACAAATGGAATCT 294
 QY 328 GCAGTATGATCTCTTGAACAGCTAAAGCTCAAAATTCGTAAGTAATTTAGTTGATG 387
 DB 295 GCAGGTTACACAGAACCAATTCAGACAGATATCAACCAAGTGTACCTTTCAAGAA 354
 QY 388 GCAATTAAGAGGTACACAGCTGATGATGACAGATGAAGCCCTATTAAGAAAGCTTTGAT 447
 DB 355 GGTCTAAAGAA-----ACATATCAAAATTGACCGATGATTTAAAAACGCTTGGGAA 408
 QY 448 GAGTACCTCCAGATGTAAGGCTCAATCATCCGCTTAATATTAAGATTAAGGCAAA 507
 DB 409 TCTTTCCATCCAGAGTGAAGCTCAAAATTCAGATGCAAGTGAAGATGATGCAAAA 468
 QY 508 GAAATTTTCGAAAAACCAAGCAGAGGTGCTGATTTTGTCAATTAAGCCAAAGATAT 567
 DB 469 GA---TGTAAGAAAGCTGCTGATTAAGGAGAGCATTTCTTAATTTAGCCAAAGCAAG 525
 QY 568 TCAAGTGAAGAAAAAAGAAAAATGTTGAGAAATTAATTAATTAATTAATTAATTAAT 627
 DB 526 TCACTGATGATGCAAGCAAGAGAGTGGGAAAAAGTGAATTTGATTCCTACTACCA 585
 QY 628 GAATACCTGAGCAAGTCAAAAAAGCCGCTTCGCTTATGATGATGATGATGATGATGAT 687
 DB 586 ACGGTTCCGGCAGAGTCAAGAGACAGCTTAAAGTGAAGATGATGATGATGATGATGAT 645
 QY 688 GTGATTTACGCAAGTGGCAGACAGCTTACAGTACCAATATTAATTTGAATCACTACT 747
 DB 646 GTGATCACTTCTTACCAATGCTTCACTTAATTAAGATTAATTAATTAATTAATTAAT 705
 QY 748 AAGAAACAGAAAAATCAATTAATTTGATGATGATGATGATGATGATGATGATGATGAT 807
 DB 706 AAAAAACAAACAAAGGCAAGCAGCATGGATTAATTAATTAATTAATTAATTAATTAAT 765

QY 808 TTGACTCAAAAACAAAATGATTTACATTTTTCACAAAGCATTTATCGGAAGAANTGCAA 867
 DB 766 ACAGACACAAAATTAAGTACAGTACCTTCAGAAATTAAGTAAAGTAAAGTAAAGTAA 825
 QY 868 GCAGCCAAATATCAAGGTTAAGACCAAGCTTCCTCAAAATATATCTTTA 913
 DB 826 GACGTAATGTAATAAATCAAAAGCAAGACCTTTGAAAAATCTTTTGA 871

RESULT 8

US-09-071-035-183
 ; Sequence 183, Application US/09071035
 ; Patent No. 6448043

GENERAL INFORMATION:

APPLICANT: Gil H. Choi
 TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
 NUMBER OF SEQUENCES: 496
 CORRESPONDENCE ADDRESS:
 ADDRESS: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA
 ZIP: 20850

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/071.035
 FILING DATE:

CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:

FILING DATE:
 ATTORNEY/AGENT INFORMATION:

NAME: A. Anders Brookes
 REGISTRATION NUMBER: 36,373
 REFERENCE/DOCKET NUMBER: P8369P2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 183:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 970 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear

US-09-071-035-183

Query Match 21.0%; Score 203.2; DB 4; Length 970;
 Best Local Similarity 55.3%; Pred. No. 1.1e-41;
 Matches 466; Conservative 0; Mismatches 359; Indels 18; Gaps 3;

QY 111 AGACCTTTCAGCATGAAGGAGTGTCAATTAAGAAACATCAATTTATGAGCAAGTGA 170
 DB 19 AGATATCCCAACATGAAGGTTCAACAAATTAATGATGATTTTATTAACCAAAATTA 78
 QY 171 AAGCAACCTTCAGGCCCAAGCTGTGTAATATATGACATCCAAAAAGTTTGAANA 230
 DB 79 AGAACAAAGCATGAGCCACAAAGGTTTATGCAAAATGTTTATTAAGTCTTTGAAGA 138
 QY 231 ACAATATGCTCAGAGCTTGTATGAATAAGAGTGTGATGATCTATTTGCCGAAGAAAAA 290
 DB 139 AAAATATGCGACAAAGTAACTGACAAAGAAATTAACAAAACTTTGACGAAGCAAGA 198
 QY 291 ACA-----ATATGCGCAAACTACCAAGTGTCTTGTACACAGCAGTATGACTCT 341
 DB 199 ACAAGTAGAACCAAGGCGGAAAGTTCTCTGATGATTAATTAACACAGCTGTTAACTGA 258
 QY 342 TGAACACGTAAAGCTCAAAATTTCTACAAAGTAATTTAGTTGAGTGGCAGTTAAGAGT 401

```

Db      259  AAAAATTCAGAAACAGTTAAACAAAGACGCTATGATGACGGTCTAA----- 312
QY      402  AGAGAGAGCTGAATTCAGAGATGACGCTATAGAAAGCCTTGATGAGTACCTCCAGA 461
Db      313  AGCCCACTTAAATTAAGATGAAAGCTTTAAAGAGCTTGGGCAAGTTCCATCCAGA 372
QY      462  TGTAAAGGCTCAATATGCTCGTCTTAATATGAAGTAAGGCCAAAGATTCGAAA 521
Db      373  AGTAGAAGCACAATATTCAGAGTTCCTTCAGAGATGATGCCA---AGCTGTCAAGAA 429
QY      522  AGCCAAAGCAGAAAGTGTGATTTTCTCAATTAGCCAAAGATTAATTCACATGAAA 581
Db      430  AGAATCAGTACGAGCGGCGGATTTCAAAAAATTTGCTAAAGAAATTCACAGATAC 489
QY      582  AACAAAGAAATATGATGGAGAAATTCCTTGTGCTTCAAGCAAGTACTGACGA 641
Db      490  TACGAAAAAAGATGGGGTAAATTAATTTGATTCACAAGCAACAGTCTCCCGA 549
QY      642  AGTCAAAAAAGCGCTTTCGCTTTAGATGATGATGCTTCTGATGATGATACGCAAC 701
Db      550  AGTTAAAGAGAGCTGCTTCAATTAAGATGCGCAAGTGTCAAGCAATTCCTCCAC 609
QY      702  TGGCACAAGAGCTTCAAGTACCAATTTACATTTGTAATCTCAGTAAAGAGAAAA 761
Db      610  AATATGCAAACTTACCAAAACCACTATCTATGATGAAATGACGAAAAACCAAA 669
QY      762  ATCATCTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 821
Db      670  AGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 729
QY      822  AATGATTCACATTTGTTCAAGCATTTTCGGAAGAAAGATTCAGAGCAAGCAATTC 881
Db      730  AGCGATTCACATTTGTTCAAGCATTTTCGGAAGATTCAGAGCAAGCAATTCAG 789
QY      882  GGTAAAGAGCAAGCTTCCAAATATCTTACCAATATATGATGATGATGATGATGAT 941
Db      790  AATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 849
QY      942  CTC 944
Db      850  CGC 852

```

RESULT 9
 US-09-107-532A-1082
 : Sequence 1082, Application US/09107532A
 : Patent No. 6583275
 :
 : GENERAL INFORMATION:
 : APPLICANT: Lynn A Doucette-Stamm and David Bush
 : TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 : ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
 :
 : NUMBER OF SEQUENCES: 7310
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: GENOME THERAPEUTICS CORPORATION
 : STREET: 100 Beaver Street
 : CITY: Waltham
 : STATE: Massachusetts
 : COUNTRY: USA
 : ZIP: 02354
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: CD-ROM ISO9660
 : COMPUTER: PC
 : OPERATING SYSTEM: <Unknown>
 : SOFTWARE: ASCII
 :
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/107,532A
 : FILING DATE: 30-Jun-1998
 :
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 60/085,598
 : FILING DATE: 14 May 1998
 : APPLICATION NUMBER: 60/051571
 : FILING DATE: July 2, 1997

```

: ATTORNEY/AGENT INFORMATION:
: NAME: Ariniello, Pamela deneke
: REGISTRATION NUMBER: 40,489
: REFERENCE/DOCKET NUMBER: GTC-012
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (781)893-5007
: TELEFAX: (781)893-8277
: INFORMATION FOR SEQ ID NO: 1082:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 909 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: circular
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Enterococcus faecium
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: (8) LOCATION 1..909
: : SEQUENCE DESCRIPTION: SEQ ID NO: 1082:
US-09-107-532A-1082

Query Match      15.7%; Score 151.2; DB 4; Length 909;
Best Local Similarity 51.3%; Pred. No. 1e-28;
Matches 407; Conservative 0; Mismatches 378; Indels 9; Gaps 2;

QY      125  TGAAGGGATGTCATTCAGAACATCTTTATGACAGAGTGAAGAACCCCTTCAG 184
Db      125  TTAAGGTGACACCTTACAGATTCAGAGATTTATGATACCCAAAGTTAAATCAAA 184
QY      185  CCCAAGAGCTGTTAAATATGACATCCAAAGATTTTGAAGAAATATGATGCTGAC 244
Db      185  ATCAACACATTTAAATTAATATGATGCTTCAACACATTTGATGATGATGATGAT 244
QY      245  AGCTTGATGATAAGAGGTGATGATGATGATGATGATGATGATGATGATGATGAT 304
Db      245  AGATTTCTTCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 304
QY      305  ACTACCAAGAGCTGCTTGTCAAGACAGATGATGATGATGATGATGATGATGATGAT 364
Db      305  ACTACCAAGAGCAATTTAAAGAGCTTGGCCAAACAGAGACATATTAAGAACTCTT 363
QY      365  GTACAGTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 424
Db      364  -----AAACGATGCTGCTTCCATATGAGCTTTGAAAGCAATGTAATTAACGATA 418
QY      425  AAGCTATAAGAAAGCCTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 484
Db      419  AAGATCTAGACACTGCTTGGAAAGAAATTTATTCAGAAAGTATGATGATGATGAT 478
QY      485  TTAATATGAAGTAAGGCCAAAGAGTTCGAAAGAGCCAAAGGCAAGAGATGCTGAT 544
Db      479  TTTGCACTAAAGAGAGAGAGATTAAGCAAAAAAAGAGCAAT---GAAGGAAATTT 535
QY      545  TTGCTCAATTTAGCCAAAGATATTCATAGATGATGATGATGATGATGATGATGATG 604
Db      536  TCTCCAAAGCTTGTTCAGAGCTTATGGAAGAAATTAATTAAGGAAAGTATGAGAAA 595
QY      605  TTACCTTTGATTCGCTTCAAGAGAGTACAGTACAGTCAAGGCTCAATTCAGCTTC 664
Db      596  TGAATTTGATTCACCAATCCAGAGATTCCTACTGAGTAAGTAAGAAAGCAGCATTTAAAC 655
QY      665  TAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 724
Db      656  TAAAGAGGAGGAGTTCTGATATCTCTGTAACAGATCCAGTCAAGGACTTATCAGCAAA 715
QY      725  AATATTAATTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 784
Db      716  GCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 775
QY      785  AAGAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 844

```

Db 776 AATCTGAGCTAGAAAAAATGCTTACCGAAGCAAAAATCTTGATACGGAATTCATGATA 835
 QY 845 GCATTATCGAAGAAATTCGAAGCAGCAATATGCAAGTAAAGACCAAGCTTCCAAA 904
 Db 836 AACAATCCGTAAATGATGAAAAAGACATGTAATTAAGATCTTATGTAAAAA 895
 QY 905 ATATCTTTACCCA 918
 Db 896 ATATATTCAAATAA 909

RESULT 10

US-08-232-463-14/C
 ; Sequence 14, Application US/08232463
 ; Patent No. 5670367

GENERAL INFORMATION:
 APPLICANT: DORNER, F.
 APPLICANT: SCHIEFLINGER, F.
 APPLICANT: FALKNER, F. G.
 TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
 NUMBER OF SEQUENCES: 52
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley & Lardner
 STREET: 1800 Diagonal Road, Suite 500
 CITY: Alexandria
 STATE: VA
 COUNTRY: USA
 ZIP: 22313-0299

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/232,463
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/07/935,313
 FILING DATE:
 APPLICATION NUMBER: EP 91 114 300.6
 FILING DATE: 26-AUG-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: BENT, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 30472/114 IMMU
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703)836-9300
 TELEFAX: (703)683-4109

TELEX: 899149
 INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 7218 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 CLONE: PT29pt-F15
 US-08-232-463-14

Query Match 6.2%; Score 59.8; DB 1; Length 7218;
 Best Local Similarity 2.7%; Pred. No. 1.5e-05;
 Matches 10; Conservative 225; Mismatches 142; Indels 0; Gaps 0;

QY 228 AAAAAATATGCTCAGAGCTTGATGATGAAGAGTGTGATGCTATGCGCAAGAAA 287
 Db 1449 AGAAGAAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1390
 QY 288 AAAACAATATGCGCAAAACACACAGTCTGCTGACACAGCAGTATGACTTTGAAC 347
 Db 1389 RRR 1330

QY 348 ACCTAAGCTCAAAATTCCTACAGTAATAATGTTAGTTGGCAGTTAAGACGTACACGA 407
 Db 1329 RRR 1270
 QY 408 AGCTGAATGACAGTGAAGCCCTATAAGAAAGCTTGTGATGACTCCAGATGTAAC 467
 Db 1269 RRR 1210
 QY 468 GGCTCAATCATCCTGCTTAATGATGACATAGCCCAAGAGTCTGCAAAAAGCCCA 527
 Db 1209 RRR 1150
 QY 528 GGCAGAAAGTGTGATTTGCTCAATTAGCCAAAGATTAATCACTGATGAAAAAANA 587
 Db 1149 RRR 1090
 QY 588 AGAAATGTGAGAAA 604
 Db 1089 RRR 1073

RESULT 11

5231168-1
 ; Patent No. 5231168
 ; APPLICANT: DIEGIEL, MORTEN; BORRE, MARTIN; JEPSEN, SOREN;
 ; VUUST, JENS; RIEBECK, KLAUS; WIND, ANNETTE; JAKOBSEN, PALLE H.
 ; TITLE OF INVENTION: MALARIA ANTIGEN
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/409,658
 ; FILING DATE: 18-SEP-1989
 ; SEQ ID NO: 1
 ; LENGTH: 3095
 ; 5231168-1

Query Match 5.4%; Score 52; DB 6; Length 3095;
 Best Local Similarity 4.4.7%; Pred. No. 0.00096;
 Matches 202; Conservative 0; Mismatches 250; Indels 0; Gaps 0;

QY 201 AATATGACCAATCCAAAAGTTTGAAGAAACATATGCTCAGAGCTGATGATAAGA 260
 Db 1089 AATATGACCAATTTTCAAGAAATTAATGAAGATGAAGAAAGTCCAGATATTCAGCATGAAT 1148
 QY 261 GGTGATGATACATAATTCGCGAAGAAAAAACAATATGCGCAAAAGTACCAAGTGTCTT 320
 Db 1149 AGTAGAAGTAGAAGAAATTAATTCAGAAAGATGATTAATAAAGTTGAACTGAAT 1208
 QY 321 GTCACAAAGCAGTATGACTCTTGAACACGTAAGCTCAATTCGTAAGTAATTAAGT 380
 Db 1209 AGTAGAAGTTGAAGAAATTTTACCAAGAAATGAAGAAAGTCAAGTCAATTAAGT 1268
 QY 381 TGAAGTGGCAGTTAAGAAAGTAGCAGAAAGTGAATGACAGATGAAGCTATTAAGAAAGC 440
 Db 1269 AGAGGTGAAGAAATTTTACAGAAAGATGATTAATAAAGAAAGTTGAACATGAATAGT 1328
 QY 441 CTTGATGATGACACTCCAGATGTAACGCTCAATATCCGCTTAATTAAGATGA 500
 Db 1329 AGAAGTTGAAGAAATTTTACAGAAAGATTAATAAAGTCAACATGAATTAAGT 1388
 QY 501 GGCCAAAGAGTCTCGAAAAAGCCAGGAGAGAGTGTGATTTGCTCAATTAAGCCAA 560
 Db 1389 GGTGAAGAAATTTCTACCAAGATTAATAAAGTTGAACATGAATTAAGT 1448
 QY 561 AGATTAATTCACGTGAGAAAAAAGAAAGAAAGTGTGAGAAATTAACCTTGAATTCG 620
 Db 1449 TGAAGAAATTTCTACCAAGATTAATAAAGTGAAGTCAACATGAATTAAGT 1508
 QY 621 TTCAACGAGAGTCCGAGCAAGTCAAAAAAG 652
 Db 1509 AGAATTTCTACCAAGAAATTAATAAAG 1540

RESULT 12


```

Db      3697 TAAACAAAAAGTGAATTAATCAATTAATTCCTAAATAAATAGCAAAATCCAC 3756
QY      768 TAAATTTGATGACTACAAATAAATAAATACTGTTATCTTGACATCAAAAACAAATGA 827
Db      3757 CAATATTAATGATAAATAAATAAATAAATAAATGATTTTAAATTTTAAACAAACAT 3816
QY      828 TTCACATTTGTTCAAAAGCATTTATCGAAAGAAATTCGACAGCACCATAATCAAG 883
Db      3817 AAAAATAATGCTCTTAAGTTTATGAACATAAAGTGTGATTAATAAATAATGATGG 3872

```

```

RESULT 14
US-09-134-001C-578
; Sequence 578, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:

```

```

; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ. ID NOS: 5674
; SEQ ID NO 578
; LENGTH: 633
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-578

```

```

Query Match      4.8%; Score 46; DB 4; Length 633;
Best Local Similarity 45.2%; Pred. No. 0.017;
Matches 169; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

```

```

QY      432 TAAGAAAGCTTTGATGATGATACCTCCAGATGTAACGGCTCAATCATCGTCTTAATAA 491
Db      225 TAAAAAAGAGCTGAGATATAGTTGAAAAATGTAACAAAGCAAAAAGAAATTTGAAAA 284
QY      492 TGAAGTAAGCCAAAGAGATCTCGAAAAAGCCAAAGCAAGAGTGTGATTTGCTCA 551
Db      285 AGAAGAGAGGCTCTGATATATTTGAAAAAGATTTAAACAAGCCAAAGCATATCTTGA 344
QY      552 ATTAGCCAAAGATTAATTCAGTGAAGAAAAACAAAAGAAATGTGAGAAATTAACCTT 611
Db      345 ACATGTAGAAAAACAAAGCAAAAGAAAAAGAGTTGAACTTGATAGCTATTAAAGA 404
QY      612 TGATCTGCTTCAACAGAGTACCTGAGCAAGTCAAAAAAGCCGCTTTGCTTTAGATGT 671
Db      405 AAAAATTAATATCATGATGCTTTATGCAAAAGGCTTAACAAAAGCACTTAATTAAGAAAA 464
QY      672 GGATGGTGTTCGTGATGATTAACAGCACTGGACACAAAGCTACAGTACAGCAATATTA 731
Db      465 AGAAGCTTTTCTTATTATTAATGAAGATATGCAACATCGGAAGTAGAGGAAAAATC 524
QY      732 CATTTAAACATCACTAAGAAAAAGAAAAATCATCTAATTAATGATGACTACAAAGAAA 791
Db      525 GAAAGATCTTTCTTAAGCATATTAAGAAATGAATTAATTAATTAATGCTTACTCAAAAGC 584
QY      792 ATTAATAACTGTTA 805
Db      585 CATTGAGAAAGTTAA 598

```

```

RESULT 15
US-09-134-001C-1407

```

```

; Sequence 1407, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al

```

```

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1407
; LENGTH: 1239
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1407

```

```

Query Match      4.7%; Score 45.2; DB 4; Length 1239;
Best Local Similarity 45.3%; Pred. No. 0.034;
Matches 164; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

```

```

QY      243 AGAGCTTGATGATGAAGAGTTGATGATGATGATGATGATGATGATGATGATGATGATG 302
Db      837 AGATTATCAAAATCAAGTCAATTAATAATGCAAGGCTTTAGCTCAAAATTAATTAAGA 896
QY      303 AAACACCAACGCTGCTTGTCAACAGCAGTATGACTTTGAACACGTAAGCTCAAT 362
Db      897 AGGCTTCAGAGTGTGATGCTGCGGCAAGACATCATTTAGTGCAGTGTGATTAAG 956
QY      363 TCGTACAAGTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 422
Db      957 TTCTATCAATATGACGCTGTAATTAATGCTGAAGACACTTGATTAAGTGAATTAACATG 1016
QY      423 TGAAGCCTTAAGAAAGCCTTTGATGATGATGATGATGATGATGATGATGATGATGATG 482
Db      1017 TATAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1076
QY      483 TCTTAATTAAGATTAAGCCAAAGAGTTCGAAAAAGCCAAAGCAGAAAGTGTGAGA 542
Db      1077 AGCAACCCAGCAGCTCAACCTGCTGTTGATGATGATGATGATGATGATGATGATGATG 1136
QY      543 TTTTGCTCAATTAGCCAAAGATTAATTCACACTGATGTAAGAAAAACAAAAGAAATGTGAGA 602
Db      1137 AATTATTAATTTAGCTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1196
QY      603 AA 604
Db      1197 GA 1198

```

```

Search completed: September 4, 2003, 21:03:46
Job time : 90 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2003, 21:02:12 : Search time 315 Seconds
(without alignments)
7053.264 Million cell updates/sec

Title: US-10-049-473a-1
Perfect score: 966
Sequence: 1 agtaacctatctcaag.....gcagtagcatcaacga 966

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1537136 seqs, 114998732 residues
Total number of hits satisfying chosen parameters: 3074272

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications_NA.*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PTCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PTCT_US_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	939	97.2	942	11	US-09-769-787-179 Sequence 179, App
2	552.4	57.2	640	9	US-09-765-272-33 Sequence 33, Appl
3	215.6	22.3	12860	10	US-09-070-927A-144 Sequence 144, Appl
4	47.6	4.9	354	13	US-10-101-487-69 Sequence 69, Appl
5	47.2	4.9	554	13	US-10-101-487-106 Sequence 106, Appl
6	46.2	4.8	522	13	US-10-101-487-71 Sequence 71, Appl
7	46.2	4.8	530	13	US-10-101-487-73 Sequence 73, Appl
8	46	4.8	6189	12	US-10-240-485-145 Sequence 145, Appl
9	46	4.8	6912	12	US-10-240-453-246 Sequence 246, Appl
10	45.6	4.7	13674	12	US-10-311-455-336 Sequence 336, Appl
11	45.6	4.7	15674	12	US-10-240-485-30 Sequence 30, Appl
12	45.4	4.7	7104	9	US-09-815-242-4580 Sequence 4580, Appl
13	45.4	4.7	7107	9	US-09-815-242-8291 Sequence 8291, Appl
14	45.4	4.7	9964	12	US-10-311-455-71 Sequence 71, Appl
15	45	4.7	2142	10	US-09-917-800A-1567 Sequence 1567, Appl
16	45	4.7	3042	10	US-09-870-759-85 Sequence 85, Appl

17	45	4.7	3042	12	US-09-751-708A-85	Sequence 85, Appl
18	45	4.7	6389	12	US-10-311-455-1934	Sequence 1934, Appl
19	45	4.7	7624	12	US-10-311-455-1953	Sequence 1953, Appl
20	44.8	4.6	17131	12	US-10-311-455-1026	Sequence 1026, Appl
21	44.4	4.6	1891	10	US-09-742-096-4	Sequence 4, Appl
22	44.4	4.6	5467	12	US-10-311-455-2082	Sequence 2082, Appl
23	44.2	4.6	225	10	US-09-764-868-1268	Sequence 1268, Appl
24	44.2	4.6	5336	12	US-10-311-455-2072	Sequence 2072, Appl
25	44.2	4.6	6685	12	US-10-311-455-879	Sequence 879, Appl
26	44.2	4.6	9686	12	US-10-311-455-1662	Sequence 1662, Appl
27	43.8	4.5	567	10	US-09-764-868-24	Sequence 24, Appl
28	43.8	4.5	567	11	US-09-955-999-31	Sequence 31, Appl
29	43.6	4.5	433	10	US-09-960-353-3277	Sequence 3277, Appl
30	43.6	4.5	6314	12	US-10-240-453-14	Sequence 14, Appl
31	43.6	4.5	8876	12	US-10-311-455-2050	Sequence 2050, Appl
32	43.4	4.5	3231	12	US-10-311-455-1262	Sequence 1262, Appl
33	43.4	4.5	3508	12	US-10-311-455-1372	Sequence 1372, Appl
34	43.4	4.5	6174	12	US-10-311-455-1482	Sequence 1482, Appl
35	43.2	4.5	7195	12	US-10-240-453-40	Sequence 40, Appl
36	43.2	4.5	7195	14	US-10-239-676-30	Sequence 30, Appl
37	43	4.5	835	9	US-09-765-272-93	Sequence 93, Appl
38	43	4.5	1714	9	US-09-765-272-3	Sequence 3, Appl
39	43	4.5	5179	12	US-10-311-455-1969	Sequence 1969, Appl
40	43	4.5	5179	12	US-10-240-453-277	Sequence 277, Appl
41	43	4.5	7255	12	US-10-311-455-1251	Sequence 1251, Appl
42	43	4.5	9483	12	US-10-311-455-349	Sequence 349, Appl
43	43	4.5	14649	12	US-10-240-453-142	Sequence 142, Appl
44	43	4.5	14649	14	US-10-239-676-122	Sequence 122, Appl
45	43	4.5	15416	12	US-10-311-455-2203	Sequence 2203, Appl

ALIGNMENTS

US-09-769-787-179	RESULT 1
Sequence 179, Application US/09769787	
Publication No. US20030091577A1	
GENERAL INFORMATION:	
APPLICANT: Microbial Technics Limited	
APPLICANT: Gilbert, Christophe	
APPLICANT: Hansbro, Philip M	
TITLE OF INVENTION: Proteins	
FILE REFERENCE: PWC/P21129NO	
CURRENT APPLICATION NUMBER: US/09/769,787	
CURRENT FILING DATE: 2001-01-26	
PRIOR APPLICATION NUMBER: GB 9816337.1	
PRIOR FILING DATE: 1998-03-27	
PRIOR APPLICATION NUMBER: US 60/125164	
PRIOR FILING DATE: 1999-03-19	
NUMBER OF SEQ ID NOS: 388	
SOFTWARE: Patentln Ver. 2.1	
SEQ ID NO 179	
LENGTH: 942	
TYPE: DNA	
ORGANISM: Streptococcus pneumoniae	
US-09-769-787-179	
Query Match	97.2% Score 939 DB 11: Length 942;
Best Local Similarity	100.0% Pred. No. 3.7e-190;
Matches 939; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
28 ATGAAGAAAAAATTTATGGCAGTGCATCAGCTATTTATCAGTACCACTTACGAGCT 87	
1 ATGAAGAAAAAATTTATGGCAGTGCATCAGCTATTTATCAGTACCACTTACGAGCT 60	
88 TGTTCAGAAAGGTGCAAGGTGCAAGCTTATCAGATGAAAGGATGTCATTACAGAA 147	
61 TGTTCAGAAAGGTGCAAGGTGCAAGCTTATCAGATGAAAGGATGTCATTACAGAA 120	
148 CACCATTTTATACGAGTGAAGCAACCCCTTCAGCCCAACAATCTGTTAAATATG 207	
121 CATCAATTTTATGACGAGTGAAGCAACCCCTTCAGCCCAACAATCTGTTAAATATG 180	

```

QY 208 ACCATCCAAAAGTTTGAACAAATATGCTCAGACCTTGATGATTAAGAGTTGAT 267
   |||||||
Db 181 ACCATCCAAAAGTTTGAACAAATATGCTCAGACCTTGATGATTAAGAGTTGAT 240
QY 268 GATGATATGCGGAGAGAAAAAACAATATGCGGAGAAACTACCAAGCTCTTGTACAA 327
   |||||||
Db 241 GATGATATGCGGAGAGAAAAAACAATATGCGGAGAAACTACCAAGCTCTTGTACAA 300
QY 328 GCAGGTATGCTCTTGAACACAGTAAAGTCAAAATTCGTAAGATTAATTAAGTTG 387
   |||||||
Db 301 GCAGGTATGCTCTTGAACACAGTAAAGTCAAAATTCGTAAGATTAATTAAGTTG 360
QY 388 GCAGGTAAAGAGTACAGAGACCTGAATGACAGATGAGCCTTAATGAAGAGCTTTGAT 447
   |||||||
Db 361 GCAGGTAAAGAGTACAGAGACCTGAATGACAGATGAGCCTTAATGAAGAGCTTTGAT 420
QY 448 GAGTACACTCCAGATGTAACGGCTCAATCATCCGCTTAATATGAGATTAAGGCCAAA 507
   |||||||
Db 421 GAGTACACTCCAGATGTAACGGCTCAATCATCCGCTTAATATGAGATTAAGGCCAAA 480
QY 508 GAGTCTCCGAAAAAGCCAGAGAGAGGCTGATTTGCTCAATTAAGCCAAAGTAAAT 567
   |||||||
Db 481 GAGTCTCCGAAAAAGCCAGAGAGAGGCTGATTTGCTCAATTAAGCCAAAGTAAAT 540
QY 568 TCAACTGATGAAAAAACAAGAAAAATGCTGAGAAATTAACCTTTGCTGCTTCAACA 627
   |||||||
Db 541 TCAACTGATGAAAAAACAAGAAAAATGCTGAGAAATTAACCTTTGCTGCTTCAACA 600
QY 628 GAAGTACCTGAGACAGTCAAAAAAGCCCTTTCGCTTTAGATGATGATGTTGTTGAT 687
   |||||||
Db 601 GAAGTACCTGAGACAGTCAAAAAAGCCCTTTCGCTTTAGATGATGATGTTGTTGAT 660
QY 688 GTGATTACAGCACTGGACACAGCCTACAGTACCAATATATACATTTGTAACACTCAT 747
   |||||||
Db 661 GTGATTACAGCACTGGACACAGCCTACAGTACCAATATATACATTTGTAACACTCAT 720
QY 748 AAGAAAAACAGAAAAATCATCTAATATTTGATGACTACAAAAGAAAAATTAAGCTGTATC 807
   |||||||
Db 721 AAGAAAAACAGAAAAATCATCTAATATTTGATGACTACAAAAGAAAAATTAAGCTGTATC 780
QY 808 TTGACATCAAAAAAACAATGATGATCAATTTGTCAAAGCTTAATGCGAAAAAGATTCGAA 867
   |||||||
Db 781 TTGACATCAAAAAAACAATGATGATGATCAATTTGTCAAAGCTTAATGCGAAAAAGATTCGAA 840
QY 868 GCAGCCAAATATCAAGGTTAAGACCAAGCCTTCCAAAATATCTTTACCCAAATATATCGGT 927
   |||||||
Db 841 GCAGCCAAATATCAAGGTTAAGACCAAGCCTTCCAAAATATCTTTACCCAAATATATCGGT 900
QY 928 GTGAGATGATCAAGCTCAAGCAGTACATCAATCAAGAA 966
   |||||||
Db 901 GTGAGATGATCAAGCTCAAGCAGTACATCAAGAA 939

```

RESULT 2

US-09-765-272-33

Sequence 33, Application US/09765272

Patent No. US20020061541

GENERAL INFORMATION:

APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

```

SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 640 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 33:
US-09-765-272-33

Query Match      57.2%; Score:552.4; DB 9; Length 640;
Best Local Similarity 94.7%; Pred. No.5,4e-108;
Matches 595; Conservative 0; Mismatches 26; Indels 7; Gaps 2;

QY 90 TTGGAAGGGGTAGAGAGTGCAGACCTTATCGCATGAAGAGGAGATGTCATTACAGACA 149
   |||||||
Db 1 TTGGAAGGGGTAGAGAGTGCAGACCTTATCGCATGAAGAGGAGATGTCATTACAGACA 60
QY 150 TCAATTTATGAGCAAGTGAAGAACCAACCCCTTACGCCCAAGTCTTGTAAATATGAC 209
   |||||||
Db 61 TCAATTTATGAGCAAGTGAAGAACCAACCCCTTACGCCCAAGTCTTGTAAATATGAC 120
QY 210 CATCCAAAAGTTTGAAGAAAAACAATATGCTCAGAGCTTGATGATTAAGAGTTGATGA 269
   |||||||
Db 121 CATCCAAAAGTTTGAAGAAAAACAATATGCTCAGAGCTTGATGATTAAGAGTTGATGA 180
QY 270 TACTATTGCCGGAAGAAAAAACAATATGCGGAAACCTACCAACGTGTCTTGTACACAGC 329
   |||||||
Db 181 TACTATTGCCGGAAGAAAAAACAATATGCGGAAACCTACCAACGTGTCTTGTACACAGC 240
QY 330 AGGTATGACTCTTGAAGAACGCTTAAGCTCAATTCGTACAGTAATTTGTTGAGTTGGC 389
   |||||||
Db 241 AGGTATGACTCTTGAAGAACGCTTAAGCTCAATTCGTACAGTAATTTGTTGAGTTGGC 300
QY 390 AGTTAAGAGGTAGCAGAGCTGAATGACAGATGAGAGCTATTAAGAAAGCCTTGTATGA 449
   |||||||
Db 301 AGTTAAGAGGTAGCAGAGCTGAATGACAGATGAGAGCTATTAAGAAAGCCTTGTATGA 360
QY 450 GTACACTCCAGATGTAACGGCTCAATCATCCGCTTAATATGAAGATTAAGGCCAAAGA 509
   |||||||
Db 361 GTACACTCCAGATGTAACGGCTCAATCATCCGCTTAATATGAAGATTAAGGCCAAAGA 420
QY 510 AGTTCTCGAAAAAGCCAGAGCGAGAGGTGCTGATTTGTCATTAAGCCAAAGATTAATTC 569
   |||||||
Db 421 AGTTCTCGAAAAAGCCAGAGCGAGAGGTGCTGATTTGTCATTAAGCCAAAGATTAATTC 480
QY 570 AACTGATGAAAAAACAAGAAAAATGTTGAGAGAAATTAACCTTTGATTCGCTTCAACAGA 629
   |||||||
Db 481 AACTGATGAAAAAACAAGAAAAATGTTGAGAGAAATTAACCTTTGATTCGCTTCAACAGA 540
QY 630 AGTACCTG--AGCAAGTCAAAAAAGCCGCTTTCGCTTTAGANGTGAATGCTGTTTGTAT 687
   |||||||
Db 541 AGTACCTGAGCAAGTCAAAAAAGCCGCTTTCGCTTTAGANGTGAATGCTGTTTGTCT 600
QY 688 G-----TGATTACGCAACTGGCACACA 710
   |||||||
Db 601 GGATGTGATTTACAGCAAACTGGGGCACACA 628

```

RESULT 3

US-09-070-927A-144

; Sequence 144, Application US/09070927A
; Patent No. US20020120116A1

GENERAL INFORMATION:

APPLICANT: Charles A. Kunsch

Patrick J. Dillon

Steven Barash

TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides

NUMBER OF SEQUENCES: 982

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 Inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/070,927A

FILING DATE: 04-May-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/046,655

FILING DATE: 1997-05-16

APPLICATION NUMBER: 60/044,031

FILING DATE: 1997-05-06

APPLICATION NUMBER: 60/066,009

FILING DATE: 1997-11-14

ATTORNEY/AGENT INFORMATION:

NAME: Kenley K. Hoover

REGISTRATION NUMBER: 40,302

REFERENCE/DOCKET NUMBER: PB369

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 144:

SEQUENCE CHARACTERISTICS:

LENGTH: 12860 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 144:

US-09-070-927A-144

Query Match 22.3%; Score 215.6; DB 10; Length 12860;

Best Local Similarity 55.4%; Pred. No. 8e-36;

Matches 513; Conservative 0; Mismatches 389; Indels 24; Gaps 4;

```
QY 28 ATGAAGAAAAATTATGCGAGGTCCATCAGTATTATCAGTACCACTTTAGCAGT 87
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4226 ATGAAGAAAAATTATGCGAGGTCCATCAGTATTATCAGTACCACTTTAGCAGT 4285
QY 88 TGTGGAAGGGTCAAGGTGACACCTTATCAGATGAAGGGGATCTATTACAGA 147
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4286 TGTGGAAGGGTCAAGGTGACACCTTATCAGATGAAGGGGATCTATTACAGA 4339
QY 148 CATCAATTTTATGACAGTGAAGCAACCTTCAGCCCAAGAGTCTTTGTAATATG 207
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4340 GATGATTTTATACCAATTTAAGACAAAGCACTAGCACAAGCGTTTGGCCAAAG 4399
QY 208 ACCATCCAAAAAGTTTGAAGAAACAATATGCTCAGAGCTTGATGATTAAGAGTTGAT 267
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4400 GTTATTTATAAAGCTTGAGAAAAAATATGCGACAAAGTAACTGACAAAGATTTCAA 4459
QY 268 GATACTATTTGCCGAGAAAAAACA-----ATATGGGGAAGAAAGTACCAAGGTGC 318
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4460 AAAAAGCTTTGACGAAGCCAAAGACAAAGTAGAAGCAAGCGGGAAGTTCTGTAGTCA 4519
```

```
QY 319 TTGTCAACAGCAGGTATGACTCTTGAACACGTAAAGCTCAAAATTCGTACAGTAAATTA 378
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4520 TTAACAACAGCTGTGTTAAGTAAACAAATTCAGAAACAGTTAAACAAAGACAGCC 4579
QY 379 GTTGAAGTGGCAGTTAAGAGGTAGCAGAGCTGAATTGACAGATGAAGCTATAGAAA 438
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4580 TATGATGACAGGTCTAAA-----AGCCCACTTAAATAATACAGATGAAGACTTAAACAA 4633
QY 439 GCCTTGTAGTACTACAGTCCAGATGATACGGCTCAAAATTCAGCTTAAATGAAGAT 498
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4634 GCTTGGGCAAGTTTCCATCCAGAGTAGAAGCAAAATTTATCCAGATGCTTCACAGAT 4693
QY 499 AAGCCCAAGAGTCTCGAAAAAGCCAAAGCAGAGGTGCTGATTTGCTCAATTAAGCC 558
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4694 GATGCCA--AAGCTGCAAGAAAAAATCACTGACGGGGGATTTCACAAAAATTTGCT 4750
QY 559 AAGATTAATTCAGTGTGATGAAAAAACAAGAAATGCTGCAAAATTAATCTTATTC 618
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4751 AAGAAAAATCAACAGATCTGCTACGAAAAAAGATGGCGGTAATAATTTGATTTCA 4810
QY 619 GCTTCAACGAGTACCTGAGCAGTCAAAAAAGCGGCTTGCTTGATGATGGATGGT 678
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4811 CAAGCAACACTGTTCTGCGAGATTAAGAGCTGCTTAAATTAAGATGCGCA 4870
QY 679 GTTCTGATGTGATTACAGCACTGACACAGCCTACAGTACCAATTTACATTTGTA 738
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4871 GTGTGAGAACCAATTTGTCGACAAATATGCAAACTCAACCAACAGCTACATAGTAG 4930
QY 739 AAGCTACTAAGAAAAAGAAAAATCATCTAATTTGATGACTACAAAGAAAAATTA 798
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4931 AAAATGACGAAAAAAGCAAAAGCAATGATGAAACCTTATGAAAAAGATGTAAG 4990
QY 799 ACTGTATCTTACTCAAAAACAAATATGTTCAATTTGCAAAAGCTTTGCAAAA 858
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4991 AAAATTTGCTGAGAAAAACAAATTTAGCCATCAAAATTTGTTGCAAGTCTTTGTCAC 5050
QY 859 GAATTCAGACAGCCCAATATCAAGGTTAAGACCAAGCCTTCCAAATATCTTTACCCA 918
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5051 GAATTAAGAGCGGCCAATGTGAATTAAGATGATGCTTCAAGAACCTTTAGCAGGC 5110
QY 919 TATATCGGTGTGGAGATTTCAAGCTC 944
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5111 TACATGCAAACTGAATCTTCAAGCGC 5136
```

RESULT 4

US-10-101-487-69

; Sequence 69, Application US/10101487

; Publication No. US20020169125A1

GENERAL INFORMATION:

APPLICANT: LEUNG, DAVID W.

APPLICANT: BERGMAN, PHILIP A.

APPLICANT: LOFOUIS, ALAN

APPLICANT: PIETZ, GREGORY E.

APPLICANT: TOMPKINS, CHRISTOPHER K.

APPLICANT: WAGGONER JR., DAVID W.

TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES

FILE REFERENCE: 077319/0329

CURRENT APPLICATION NUMBER: US/10/101,487

PRIOR FILING DATE: 2002-03-20

PRIOR APPLICATION NUMBER: 60/277,705

NUMBER OF SEQ ID NOS: 116

SOFTWARE: Patent In Ver. 2.1

SEQ ID NO 69

LENGTH: 554

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Synthetic

OTHER INFORMATION: oligonucleotide

FEATURE:

NAME/KEY: CDS
LOCATION: (1)..(528)
US-10-101-487-69

Query Match 4.9%; Score 47.6; DB 13; Length 554;
Best Local Similarity 44.3%; Pred. No. 1.2;
Matches 194; Conservative 0; Mismatches 244; Indels 0; Gaps 0;

QY 215 AAAAGTTTGAACAAATATGCTCAGACCTTGAATGAAGAGTGTGATACCA 274
DB 2 ATAAATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 61
QY 275 TTGCGGAG 334
DB 62 AGGAG 121
QY 335 TGACTCTTGAACACGTAAAGCTCAATTCGTACAAATTAATTAAGTTGCACTTA 394
DB 122 AGGAG 181
QY 395 AGAAGTAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 454
DB 182 AGGAG 241
QY 455 CTCGAGAGTAAAGGCTCAATTCCTTAAATTAAGAGAGAGAGAGAGAG 514
DB 242 AGGAG 301
QY 515 TCGAG 574
DB 302 AGGAG 361
QY 575 ATGAG 634
DB 362 AGGAG 421
QY 635 CTGAGCAAGTCAAAAAG 652
DB 422 AGGAG 439

RESULT 5

US-10-101-487-106
Sequence 106, Application US/10101487
Publication No. US20020169125A1
GENERAL INFORMATION:
APPLICANT: LEUNG, DAVID W.
APPLICANT: BERGMAN, PHILIP A.
APPLICANT: LOFOULIST, ALAN
APPLICANT: PIETZ, GREGORY E.
APPLICANT: TOMPKINS, CHRISTOPHER K.
APPLICANT: WAGGONER JR., DAVID W.
TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
FILE REFERENCE: 077319/0329
CURRENT APPLICATION NUMBER: US/10/101,487
CURRENT FILING DATE: 2002-03-20
PRIOR APPLICATION NUMBER: 60/277,705
NUMBER OF SEQ ID NOS: 116
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 106
LENGTH: 554
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
OTHER INFORMATION: oligonucleotide
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(537)
US-10-101-487-106

Query Match 4.9%; Score 47.2; DB 13; Length 554;
Best Local Similarity 44.4%; Pred. No. 1.5;
Matches 190; Conservative 0; Mismatches 238; Indels 0; Gaps 0;

QY 225 TGAAG 284
DB 12 TGAAG 71
QY 285 AAAAG 344
DB 72 AGGAG 131
QY 345 AACAGTAAAGCTCAATTCCTTAAATTAAGAGAGAGAGAGAGAGAGAG 404
DB 132 AGAAG 191
QY 405 AGAAGTAAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 464
DB 192 AGAAG 251
QY 465 AACGCTCAATTCATCCCTTAAATTAAGAGAGAGAGAGAGAGAGAGAG 524
DB 252 AGAAG 311
QY 525 CAGGAG 584
DB 312 AGAAG 371
QY 585 AAAAG 644
DB 372 AGAAG 431
QY 645 CAAAAAG 652
DB 432 AGAAG 439

RESULT 6

US-10-101-487-71
Sequence 71, Application US/10101487
Publication No. US20020169125A1
GENERAL INFORMATION:
APPLICANT: LEUNG, DAVID W.
APPLICANT: BERGMAN, PHILIP A.
APPLICANT: LOFOULIST, ALAN
APPLICANT: PIETZ, GREGORY E.
APPLICANT: TOMPKINS, CHRISTOPHER K.
APPLICANT: WAGGONER JR., DAVID W.
TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
FILE REFERENCE: 077319/0329
CURRENT APPLICATION NUMBER: US/10/101,487
CURRENT FILING DATE: 2002-03-20
PRIOR APPLICATION NUMBER: 60/277,705
NUMBER OF SEQ ID NOS: 116
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 71
LENGTH: 522
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
OTHER INFORMATION: oligonucleotide
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(522)
US-10-101-487-71

Query Match 4.8%; Score 46.2; DB 13; Length 522;
Best Local Similarity 44.3%; Pred. No. 2.3; Mismatches 238; Indels 0; Gaps 0;
Matches 189; Conservative 0;


```

Oy      872  CCAATATCAAGGTTAAGGACCAAGCCTTCCAAAATATCTTTACCCAATAT  921
        || | | | | | | | | | | | | | | | | | | | | | | | | |
Db      3238 ACATTTCACATACACATACCTTATACCTTAACACACATTTCTTCTCAATAT  3189

```

RESULT9
US-10-240-453-246/c
; Sequence 246, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:

OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-246

Query Match	Similarity	Score	DB	Length
Best Local	50.5%	Pred. No. 7.4		
Matches	112; Conservative	0; Mismatches	110; Indels	0; Gaps
QY	699	AACTGGCAGACAGGCTACAGTACGCAATATTACATGTAAACTACTAGAAACAGA	758	
Db	1573	AACTTAACCTCTAAATAAAAAACTCAAAAAAAACCTTATTAACCAACAAATACCAATATAA	1514	
QY	759	AAATATCACTATATTGTGATGACACAAAGAAAATTAATAAAGCTTATCTTGACACAA	818	
Db	1513	AAAAACAAAAAAACGATTAATATTTAAATATAAAAAATTTAAACTTAACT	1455	
QY	819	ACAAATGATTCACATTTGGTCAAAAGCATTAATCGAAAGAAATTCGAAGACCCAAATAT	878	
Db	1453	AAAAATATATCTTAATTAATTTTCATTAATAAATTAATTAATCACTAATACACACTAACAACCT	1394	
QY	879	CAAGTTAAGGACCAAGCCTCCAAATATCTTTACCCAAATA	920	
Db	1393	TATATATTAATTAACCAATATCAATATTCATTAACCAACTATAA	1352	

```

: RESULT 10
: US-10-311-455-336/C
: Sequence 336, Application US/10311455
: Publication No. US20030143606A1
: GENERAL INFORMATION:
: APPLICANT: OLEK, Alexander
: APPLICANT: PIEPENBROCK, Christian
: APPLICANT: BERLIN, Kurt
: TITLE OF INVENTION: Diagnosis of diseases
: TITLE OF INVENTION: Cytosine methylation
: FILE REFERENCE: 5013.1014
: CURRENT APPLICATION NUMBER: US/10/311,455

```

```

? CURRENT FILING DATE: 2002-12-16
? PRIOR APPLICATION NUMBER: PCT/EP01/07537
? PRIOR FILING DATE: 2001-07-02
? PRIOR APPLICATION NUMBER: DE 100332529.7
? PRIOR FILING DATE: 2000-06-30
? PRIOR APPLICATION NUMBER: DE 10043826.1
? PRIOR FILING DATE: 2000-09-01
? NUMBER OF SEQ ID NOS: 2424
? SEQ ID NO 336
? LENGTH: 15674
? TYPE: DNA
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-336

```

Query Match	4.7%;	Score 45.6;	DB 12;	Length 15674;
Best Local Similarity	45.7%;	Pred. No. 13;		
Matches 159;	Conservative 0;	Mismatches 189;	Indels 0;	Gaps 0

[illegible]

```

; RESULT 11
; US-10-240-485-30/C
; Sequence 30, Application US/10240485
; Publication No. US20030148327A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: Metastasis
; FILE REFERENCE: 5013.1007
; CURRENT APPLICATION NUMBER: US/10/240,485
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03970
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10033529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 202
; SEQ ID NO 30
; LENGTH: 15674
; TYPE: DNA
; ORGANISM: Artificial Sequence

```

US-09-815-242-4580

336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

326 AAGCAGTATGACTCTTGAACACGTAAGCTCAAATTCGTACAAGTAAATTAGTTGAGT 385

Db	6722	AAGGACCTTCGACCTGAGGATAGTCAAAAAGCTGACATGGCAACGACATATCAAAAAGTGA	6781
Qy	386	TGGCAGCTTAAGAAAGGTAGCAGAACTGTAATTGACAGATGAAGCCTATATGAAGAGCCTTTG	445
Db	6782	ATCAAGCATCAATTGGTGTGCTCAACTGCGAGTGTAAACAATAAAGCTACGCAAAACGCGGTG	6841
Qy	446	ATGAGTACACTCCAGATGTAAAGGCTCAAAATCATCCGTCCTTAATTAATGAATGAAGGCA	505
Db	6842	CCAAATGCATCTCCAGCTCCAGCTTTCACAAAGGAGCAATAGCTAATCAATATATGCTTGA	6901
Qy	506	AAGAGTTCCTCGAAAAAGCCAAAGCAGAGAAGTGGATTTGGTCATATAGCAAAAGATA	565
Db	6902	ATGTAACTAAAGACTTAAGAAAAACAAGCAAAATGCAAATTCAGCCCAACAAGAAAAAGTTA	6961
Qy	566	ATTCACCTGATGAAAAAACAATA	588
Db	6962	ATTAAGCCTTAACACAAACAGCTAAA	6984

```

RESULT 14
; US-10-311-455-71/C
; Sequence, 71, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: Cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 71
; LENGTH: 9964
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-71

```

[illegible]

```

Db      8831  AAACATAAAAAAATCAATTAATCTAAAAATTTATACAAAAATTCATAAAAATPAAAAAC  8772Z
QY      499   AAGGCCAAGAAGTTCCTCGAAAAAGCCAAGGCAGAGGCTGCTGATTTTGTCTAATTAGCC  558
Db      8771  AAAAATAAAAAAATCCACAAAAATATAAAATCA-----ATCAACACCAAAACATNA  8720Z
QY      559  AAAGATAATCACTGATGAAAAAACAAGAAAAATGSGAGAAATTAACCTTGATCT  618
Db      8719  AAAAAATCTACCAAAAAATTCAAAAAATAATTAATTTAAAAAATAAAATCATACCCCT  8660Z
QY      619  GCTTCAACAGAGTACTGAGCACAAGTCAAAAAAGCCGCTTGCGTTAGATGTGGATGCT  678
Db      8659  AATTAAAAAAATTAAAAAAAACTAAAAAAACAAAAAATAAAAAATTAACAA  8600Z
QY      679  GTTTCGTGATGATTACAGCACTGGCACAGACGCTACAGTAGCCATATTACATTGTA  738
Db      8599  AATCAAAAAATTAATAAACAAAAAATTAACATAACAAAAAATTAATAATTAAAAAATTA  8540Z
QY      739  AAACCTACTAAGAAAAACGAAAAAATCATCTAATTGATGCTCACAAGAAAAATTAAA  798
Db      8539  AAACCTAAAAAAACACCAAAAAAACAACATTAATAAAAAAATAAAAAAATAAATATTA  8480Z
QY      799  ACTGTTATCTGACCTCAAAAAACAATAATGATTCACACATTTGTTCAAGCATATGGA  858
Db      8479  AAAAAAAAAACAATTTAAAAAACAATAATCAAAAATTAACAAAAATTAATAACAATTA  8420Z
QY      859  GAATTGCAGACCAATATCAAGGTTAAGACCAAGCCTTCCAAATATCTTTACCCA  918
Db      8419  ATAAAAAAAACATAAAAAAATATTAATAAACTTAAACACTAACCTATTAATAAAATTA  8360Z
QY      919  TATAT 923
Db      8359  TPAAT 8355

```

```

RESULT 15
; Sequence 1567, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION
; APPLICANT: Mendicor, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1567
; LENGTH: 2142

```


BEST AVAILABLE COPY

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2003, 19:14:57 : Search time 2071 Seconds

(without alignments)
11336.612 Million cell updates/sec

Title: US-10-049-473a-1

Perfect score: 966

Sequence: 1 agtaacctatctcaag.....gcagtagatcatcaagaa 966

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 segs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_estch:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	72.4	7.5	868	28	BH161951
C 2	71.6	7.4	843	28	AZ551618
C 3	71.6	7.4	931	28	BH160272
C 4	71	7.3	877	28	AZ531291

Result	Score	Query Match	Length	DB ID	Description
C 5	71	7.3	908	28	AZ548467
C 6	71	7.3	912	28	AZ551092
C 7	69.8	7.2	905	28	AZ550256
C 8	66.6	6.9	849	28	AZ546009
C 9	66	6.8	938	28	AZ676392
C 10	64.8	6.7	890	28	AZ530768
C 11	64.6	6.7	976	28	BH149983
C 12	64.2	6.6	880	28	AZ669474
C 13	63.2	6.5	665	13	BH432980
C 14	62.8	6.5	665	13	AZ529191
C 15	62.8	6.5	861	28	AZ529191
C 16	61.6	6.4	880	13	BH266138
C 17	60.8	6.4	677	12	BH161747
C 18	60.4	6.3	532	12	CNS0039G
C 19	60.4	6.3	719	14	CA855416
C 20	60.2	6.2	816	28	AZ535744
C 21	60	6.2	890	28	BH146886
C 22	59.8	6.2	879	28	AZ550718
C 23	59.8	6.2	942	28	BH148582
C 24	59.2	6.1	656	13	BH144040
C 25	59.2	6.1	721	13	BH141179
C 26	59.2	6.1	891	28	AZ683582
C 27	59.2	6.1	942	14	CD050127
C 28	59	6.1	906	28	BH153606
C 29	58.8	6.1	900	28	AZ549980
C 30	57.4	5.9	748	29	BX137558
C 31	57	5.9	671	13	BH273704
C 32	57	5.9	727	29	BX159538
C 33	57	5.9	1061	14	CD386766
C 34	56.8	5.9	1184	28	BH13117
C 35	56.6	5.9	959	29	CNS00655
C 36	56.4	5.8	871	13	BH262815
C 37	56	5.8	888	28	AZ528430
C 38	55.8	5.8	481	13	BH268570
C 39	55.8	5.8	1101	29	CNS0039R
C 40	55.4	5.7	1152	29	BZ696911
C 41	55.2	5.7	644	13	BH136514
C 42	55.2	5.7	916	28	AZ671886
C 43	55	5.7	742	13	BX462820
C 44	54.8	5.7	452	13	BH138725
C 45	54.8	5.7	770	12	BH169947

ALIGNMENTS

RESULT 1
BH161951/c 868 bp DNA linear GSS 24-SEP-2001
LOCUS ENTBR12ZR Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic, genomic survey sequence.
ACCESSION BH161951
VERSION BH161951.1 GI:15735389
KEYWORDS GSS.
SOURCE Entamoeba histolytica
ORGANISM Entamoeba histolytica
REFERENCE 1 (bases 1 to 868)
AUTHORS Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library (2001)
COMMENT Unpublished
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun

OY 518 AAAAGCCAGCAGAGAGTCTGATTTTCTCAATTAGCCAAAGATTAATTCAGTATG 577
 DB 309 AAGGAAGATGATGAAGAGAGATGATGAAGAGATGATGAAGAGAGATGATG 250
 OY 578 AAAAAGCAAGAAATGATGAGAAATTTACCTTTGATTTCTTCAACAGAGTACTG 637
 DB 249 ATGAAGACAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 190
 OY 638 AGCAAGTCAAAAAGCCCTTTGCTTATGATGATGATGATGATGATGATGATGATG 697
 DB 189 AAGGAAGATGATGAAGAGAGATGATGATGATGATGATGATGATGATGATGATG 130
 OY 698 CAATGCGACACAGCCTTACAGTACCAATATTACATTTGTAACCTCACTAAGAAACAG 757
 DB 129 ATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATGAAG 70
 OY 758 AAAAATCATCTAATTTGATGCTTCAAAAGAAATTAATAA 799
 DB 69 AAGATGATGAAGATGATGAAGAGATGATGAAGATGATGAAGATGATGAAGATG 28

RESULT 3

BHI60272 931 bp DNA linear GSS 24-SEP-2001
 LOCUS ENT0049TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
 DEFINITION genomic, genomic survey sequence.

ACCESSION BHI60272
 VERSION BHI60272.1 GI:15733710

KEYWORDS GSS.
 SOURCE Entamoeba histolytica
 ORGANISM Entamoeba histolytica
 Eukaryota; Entamoebidae; Entamoeba.

REFERENCE 1 (bases 1 to 931)
 AUTHORS Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.
 TITLE Determination of clone end sequences from Entamoeba histolytica
 HMI:IMSS sheared DNA library (2001)
 JOURNAL Unpublished

COMMENT Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b.loftus@tigr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
 DNA library
 Seq primer: M13-Reverse

Class: Shotgun
 High quality sequence start: 8
 High quality sequence stop: 677.

FEATURES
 Source Location/Qualifiers

1..931

/organism="Entamoeba histolytica"

/mol_type="genomic DNA"

/strain="HMI:IMSS"

/db_xref="taxon:5759"

/clone_lib="Entamoeba histolytica Sheared DNA"

/note="vector: pROSL; Site_1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + 1 method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing projects. In Genome
 Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Barrell, Oxford University Press, 1999)."
 402 a 51 c 279 g 199 t

BASE COUNT
 ORIGIN

Query Match 7.4%; Score 71.6; DB 28; Length 931;
 Best Local Similarity 45.2%; Pred. No. 0.068;
 Matches 263; Conservative 0; Mismatches 319; Indels 0; Gaps 0;

OY 218 AAGTTTGAAGAAACAAATGCTTCAAGCTTGTATGAAGAGTGTATGATGATG 277
 DB 96 ATGATGATGATGAAGAGATGATGATGATGATGATGATGATGATGATGATGATG 155
 OY 278 CCGAAGAAAAAAGCAATATGATGAGAAACTACCAAGCTGTCTTGCACAGCAGGTATGA 337
 DB 156 AAGGAAGATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATGAAGAGATG 215
 OY 338 CTCTTGAAGACAGTAAAGCTCAATTCGTACAAATTAATTAAGTGGCAGTTAAGA 397
 DB 216 ATGATGATGAAGATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATGAAG 275
 OY 398 AGTGAAGAGATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATGAAGAGATG 457
 DB 276 AAGGAAGAGATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATGAAGAGATG 335
 OY 458 CAGATGTAACGGCTCAATTCATCCGCTTAAATTAATGAAGATGAAGAGATGAAG 517
 DB 336 AAGATGAAGATGATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATGAAG 395
 OY 518 AAAAAGCCAGCAGAGAGTCTGATTTTCTCAATTAGCCAAAGATTAATTCAGTATG 577
 DB 396 AAGGAAGATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATG 455
 OY 578 AAAAAGCAAGAAATGATGAGAAATTTACCTTTGATTTCTTCAACAGAGTACTG 637
 DB 456 ATGAAGACAGCAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 515
 OY 638 AGCAAGTCAAAAAGCCCTTTGCTTATGATGATGATGATGATGATGATGATGATG 697
 DB 516 AAGGAAGATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATG 575
 OY 698 CAATGCGACACAGCCTTACAGTACCAATATTACATTTGTAACCTCACTAAGAAACAG 757
 DB 576 ATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATGAAGATG 635
 OY 758 AAAAATCATCTAATTTGATGCTTCAAAAGAAATTAATAA 799
 DB 636 AAGATGATGAAGATGATGAAGAGATGATGAAGAGATGATGAAGATGATGAAGATG 677

RESULT 4

A2531291/c 877 bp DNA linear GSS 03-NOV-2000
 LOCUS ENT0034TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
 DEFINITION genomic, genomic survey sequence.

ACCESSION A2531291
 VERSION A2531291.1 GI:11085838

KEYWORDS GSS.
 SOURCE Entamoeba histolytica
 ORGANISM Entamoeba histolytica
 Eukaryota; Entamoebidae; Entamoeba.

REFERENCE 1 (bases 1 to 877)
 AUTHORS Loftus, B., Van Aken, S. and Fraser, C.
 TITLE Determination of clone end sequences from Entamoeba histolytica
 HMI:IMSS sheared DNA library
 JOURNAL Unpublished

COMMENT Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b.loftus@tigr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
 DNA library
 Seq primer: M13-Reverse
 Class: Shotgun
 High quality sequence start: 22

High quality sequence stop: 829.

FEATURES
Location/Qualifiers
1..877

/organism="Entamoeba histolytica"
/mol_type="genomic DNA"
/strain="HMI:IMSS"

/db_xref="taxon:5759"

/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHO51; Site 1; Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + 1 method used for the library construction is described in detail in Smith, H.O., and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999)."

Query Match
Best Local Similarity 46.9%; Pred. No. 0.086; Length 877;
Matches 221; Conservative 0; Mismatches 250; Indels 0; Gaps 0;

BASE COUNT
Origin 156 a 268 c 29 g 424 t

7.3%; Score 71; DB 28; Length 877;

Query Match
Best Local Similarity 46.9%; Pred. No. 0.086; Length 877;
Matches 221; Conservative 0; Mismatches 250; Indels 0; Gaps 0;

218 AGCTTTTGAACAAATATGCTCAGAGCTGATGATGAAGAGTTGATGATCTATTG 277

682 AAGAGATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATGAAGAGATG 623

278 CCGAAGAAAAAACAATATGCGGAAACCTACAGCTGCTGTCACAGCAGGTATGA 337

622 AGCATGAGACGACGATATGATTAATGAGATGATGATGAAGAGAGATGATG 563

338 CTCTTGAACACGTAAAGCTCAATTCGTACAGTAATTAATGATGATGATGATG 397

562 ATGAGATGATGAAGAGATGATGAAGAGATGATGATGAAGAGATGATGATG 503

398 AGGTACGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 457

502 AAGAGATGATGAAGAGATGATGAAGAGATGATGATGAAGAGATGATGATG 443

458 CAGATGTAAGGCTCAATATCCCTTAATTAATGAAGATGAAGAGAGATGATG 517

442 ATGAGAGATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATGATG 383

518 AAAAGCCAGAGAGATGATGATGATGATGATGATGATGATGATGATGATG 577

382 AAGATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATGATGATGATG 323

578 AAAAAACAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATG 637

322 AAGATGATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATGATGATG 263

638 AGCAAGTCAAAAAAGCCGCTTTCGTTAGATGATGATGATGATGATGATGATG 688

262 ATGATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATGATGATGATG 212

RESULT 5
A2548467

LOCUS A2548467 908 bp DNA linear GSS 14-NOV-2000

DEFINITION Entamoeba histolytica Sheared DNA Entamoeba histolytica

ACCESSION A2548467

VERSION A2548467.1 GI:11172102

KEYWORDS GSS.

SOURCE Entamoeba histolytica

ORGANISM Entamoeba histolytica

REFERENCE 1 (bases 1 to 908)

AUTHORS

Loftus, B., Van Aken, S. and Fraser, C.

JOURNAL
TITLE
HMI:IMSS Sheared DNA Library

COMMENT
Unpublished

Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543

Email: b1loftus@tigr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library

Seq primer: M13-Reverse

Class: shotgun
High quality sequence start: 17
High quality sequence stop: 828.

Location/Qualifiers
1..908

/organism="Entamoeba histolytica"
/mol_type="genomic DNA"
/strain="HMI:IMSS"

/db_xref="taxon:5759"

/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHO51; Site 1; Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + 1 method used for the library construction is described in detail in Smith, H.O., and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999)."

Query Match
Best Local Similarity 46.9%; Pred. No. 0.085; Length 908;
Matches 221; Conservative 0; Mismatches 250; Indels 0; Gaps 0;

218 AAGTTTGAACAAATATGCTCAGAGCTGATGATGAAGAGTTGATGATCTATTG 277

98 AAGAGATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATGAAGAGATG 157

278 CCGAAGAAAAAACAATATGCGGAAACCTACAGCTGCTGTCACAGCAGGTATGA 337

158 AAGATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATG 217

338 CTCTTGAACACGTAAAGCTCAATTCGTACAGTAATTAATGATGATGATGATG 397

218 ATGAGAGATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATGATGATG 277

398 AGGTACGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 277

278 AAGAGAGATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATGATGATG 337

458 CAGATGTAAGGCTCAATATCCCTTAATTAATGAAGATGAAGAGAGATGATG 517

338 ATGAGAGATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATGATGATG 397

518 AAAAGCCAGAGAGATGATGATGATGATGATGATGATGATGATGATGATG 577

398 AAGATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATGATGATGATG 457

578 AAAAAACAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATG 637

458 AAGATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATGATGATGATG 517

638 AGCAAGTCAAAAAAGCCGCTTTCGTTAGATGATGATGATGATGATGATGATG 688

Db 518 ATGATGATGAAGAAGATGACATGATGATGATGAAGAAGATGATGATG 568

RESULT 6
LOCUS A2551092
DEFINITION A2551092 912 bp DNA linear GSS 14-NOV-2000
ENTRFJ22TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, genomic survey sequence.
A2551092
A2551092.1 GI:11176393
GSS.
Entamoeba histolytica
Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
1 (bases 1 to 912)
Loftus,B., Van Aken,S. and Fraser,C.
Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library
Unpublished
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 17
High quality sequence stop: 861.
Location/Qualifiers

```

BASE COUNT      423 a
ORIGIN           34 c      264 g      191 t

/organism="Entamoeba histolytica"
/mol_type="genomic DNA"
/strain="HMI: IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: pHOSt1; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (< 2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaubin and B.
Bartell, Oxford University Press, 1999)."

```

	Query Match	7.3% Best Local Similarity	Score 71: 46.98%	DB 28: Pred. No. 0.085	Length 912: Matches 221:	Mismatches 0:	Indels 250:	Gaps 0:
Qy	218	AACTTTTGA	AAAAACAATATG	GGCTCAGACGCTTGATG	ATAAAGAGTGTGATAC	TATG	277	
Db	316	AAGAAGATG	ATGAAGAAGATG	ATGAAGAAGAAGATG	ATGAAGATGAAGATG	ATGAAG	375	
Qy	278	CCGAAGAAAA	AAACAATATG	GGCGAAAC	TACCAACGTC	TCTTGTC	CAACACGATATGA	337
Db	376	ACGATGAAG	ACGACGAATATGA	ATTGAAATG	ATGATGAAGAA	AAAGAAAGATG	ATG	435
Qy	338	CTCTTGA	AAACGCTAAAGCTCA	AAATTCCTACAAAGTA	ATTAATGTTGAGCTG	CGACAGTTAGA	397	
Db	436	ATGAAGATG	ATGAAGACGATGA	AGCAGATGAAGCAGCAATG	ATGATTTGAAGATATG	ATGATGATG	495	
Qy	398	AGTGAACGA	AGCTGAATTGACAG	TGAAGCCCTATGA	AAAGCCCTTGATGA	GTACACTC	457	

Db	496	AAACAAGATGATGATGATGAAGAAGAAGATGATGATGAAGATGATGATGAAGAAG	555
Qy	458	CAGATGTAAAGGCTCAATCATCCGCTTTAATGAAGATAAAGCCAAAGAAGTCTCG	517
Db	556	ATTAAACAAGATGATGAAGAAGATGATGAAGATGATGAAGAAGAAGAAGATGATG	615
Qy	518	AAAAAGCCAGGCGAAGGCTGCTGATTTTGCATTAAGCCAAAGATTAATCAACTGATG	577
Db	616	AAGATGATGAAGCGATGAAGACGACGAATGATTAATGAAGATGATGAAGAAGAAAG	675
Qy	578	AAAAAACAAAGAAATGTGTGAAGAAATTAACCTTGATTCGTTCAACGAAGTAACCTG	637
Db	676	AAATGATGATGTAAGAAGAAGAAGATGATGATGAAGATGATGATGAAGAAGAAGATGAAG	735
Qy	638	AGCAAGTCAAAAAGCCCGCTTTCGCTTTAGATGTGATGTGATGGTGTTCGATG	688
Db	736	ATATATATGAAGAAGAAGATGAAGATGATGATGAAGAAGAAGAAGATGATGATG	786

FEATURES	REFERENCE	DEFINITION	LOCUS	RESULT 7
Location/Qualifiers	Loftus, B., Van Aken, S. and Fraser, C.	Entamoeba histolytica	A2550256/c	1
Class: shotgun	Determination of clone end sequences from Entamoeba histolytica	Entamoeba histolytica	A2550256	2
Seq primer: M13-Reverse	HMI:IMSS sheared DNA library	Entamoeba histolytica	A2550256	3
High quality sequence start: 100	Unpublished	Entamoeba histolytica	A2550256	4
High quality sequence stop: 872.	Contact: Brendan J Loftus	Entamoeba histolytica	A2550256	5
	Department of Eukaryotic Genomics	Entamoeba histolytica	A2550256	6
	The Institute for Genomic Research	Entamoeba histolytica	A2550256	7
	9712 Medical Center Dr., Rockville, MD 20850, USA	Entamoeba histolytica	A2550256	8
	Tel: 301 838 0208	Entamoeba histolytica	A2550256	9
	Fax: 301 838 3543	Entamoeba histolytica	A2550256	10
	Email: bjloftus@igf.org	Entamoeba histolytica	A2550256	11
	Clones are derived from the Entamoeba histolytica HMI:IMSS sheared	Entamoeba histolytica	A2550256	12
	DNA library	Entamoeba histolytica	A2550256	13
		Entamoeba histolytica	A2550256	14
		Entamoeba histolytica	A2550256	15
		Entamoeba histolytica	A2550256	16
		Entamoeba histolytica	A2550256	17
		Entamoeba histolytica	A2550256	18
		Entamoeba histolytica	A2550256	19
		Entamoeba histolytica	A2550256	20
		Entamoeba histolytica	A2550256	21
		Entamoeba histolytica	A2550256	22
		Entamoeba histolytica	A2550256	23
		Entamoeba histolytica	A2550256	24
		Entamoeba histolytica	A2550256	25
		Entamoeba histolytica	A2550256	26
		Entamoeba histolytica	A2550256	27
		Entamoeba histolytica	A2550256	28
		Entamoeba histolytica	A2550256	29
		Entamoeba histolytica	A2550256	30
		Entamoeba histolytica	A2550256	31
		Entamoeba histolytica	A2550256	32
		Entamoeba histolytica	A2550256	33
		Entamoeba histolytica	A2550256	34
		Entamoeba histolytica	A2550256	35
		Entamoeba histolytica	A2550256	36
		Entamoeba histolytica	A2550256	37
		Entamoeba histolytica	A2550256	38
		Entamoeba histolytica	A2550256	39
		Entamoeba histolytica	A2550256	40
		Entamoeba histolytica	A2550256	41
		Entamoeba histolytica	A2550256	42
		Entamoeba histolytica	A2550256	43
		Entamoeba histolytica	A2550256	44
		Entamoeba histolytica	A2550256	45
		Entamoeba histolytica	A2550256	46
		Entamoeba histolytica	A2550256	47
		Entamoeba histolytica	A2550256	48
		Entamoeba histolytica	A2550256	49
		Entamoeba histolytica	A2550256	50
		Entamoeba histolytica	A2550256	51
		Entamoeba histolytica	A2550256	52
		Entamoeba histolytica	A2550256	53
		Entamoeba histolytica	A2550256	54
		Entamoeba histolytica	A2550256	55
		Entamoeba histolytica	A2550256	56
		Entamoeba histolytica	A2550256	57
		Entamoeba histolytica	A2550256	58
		Entamoeba histolytica	A2550256	59
		Entamoeba histolytica	A2550256	60
		Entamoeba histolytica	A2550256	61
		Entamoeba histolytica	A2550256	62
		Entamoeba histolytica	A2550256	63
		Entamoeba histolytica	A2550256	64
		Entamoeba histolytica	A2550256	65
		Entamoeba histolytica	A2550256	66
		Entamoeba histolytica	A2550256	67
		Entamoeba histolytica	A2550256	68
		Entamoeba histolytica	A2550256	69
		Entamoeba histolytica	A2550256	70
		Entamoeba histolytica	A2550256	71
		Entamoeba histolytica	A2550256	72
		Entamoeba histolytica	A2550256	73
		Entamoeba histolytica	A2550256	74
		Entamoeba histolytica	A2550256	75
		Entamoeba histolytica	A2550256	76
		Entamoeba histolytica	A2550256	77

```

Query Match          7.2%, Score 69.8, DB 28, Length 905;

BASE COUNT
ORIGIN
203 a      260 c      46 g      396 t

/organism="Entamoeba histolytica"
/mol_type="genomic DNA"
/strain="HML:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notice="Vector: pHOSt1; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification.. Exp. Parasitol.
77:450-). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaubin and B.
Barell, Oxford University Press, 1999)."
```


ACCESSION A2676392
 VERSION A2676392.1 GI:11813538
 KEYWORDS GSS.
 SOURCE Entamoeba histolytica
 ORGANISM Entamoeba histolytica
 REFERENCE Eukaryota; Entamoebidae; Entamoeba.
 1 (bases 1 to 938)
 AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
 TITLE Determination of clone end sequences from Entamoeba histolytica
 HML:IMSS sheared DNA library
 JOURNAL Unpublished
 COMMENT Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b.loftus@tigr.org
 Clones are derived from the Entamoeba histolytica HML:IMSS sheared
 DNA library
 Seq primer: M13-Reverse
 Class: shotgun
 High quality sequence start: 34
 High quality sequence stop: 809.
 Location/Qualifiers
 1..938
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HML:IMSS"
 /db_xref="taxon:5759"
 /note="Vector: PHOS1; Site 1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing. projects. In Genome
 Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Barrell, Oxford University Press, 1999)."
 BASE COUNT 132 a 201 c 107 g 498 t
 ORIGIN
 Query Match 6.8%; Score: 66; DB 28; Length 938;
 Best Local Similarity 46.7%; Pred. No. 0.53;
 Matches 210; Conservative 0; Mismatches 240; Indels 0; Gaps 0;
 QY 353 AACCTGAANTTCGTACAGTAATTTAGTTGGCAGTTAAGAGCTAGCAAGCTG 412
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 792 AACCTGAAGCAAAAGCTAAAGAGAGCTAGCAAAAGCTAAAGAGAGCTGAACAA 733
 QY 413 AATTGACAGTAGAGCTATAGAAAGCCCTTGTAGTAGTACCTCCAGATGTAAAGCGCTC 472
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 732 AACCTGAAGAGAGAGCTAAAGAGAGCTGAAGAGAGCTGAAGAGAGAGAGAG 673
 QY 473 AAATCATCGCTTAAATATGAAGATAGAGCCCAAGAGTTCTCGAAAAAGCCAAAGCAG 532
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 672 AACCTGAAGCAAAAGCTAAAGAGAGCTGAAGCAAAAGCCAAAGAGAGAGAGCA 613
 QY 533 AAGCTGCTGATTTTGCATTTAGCCAAAGATTAATCACTGATGAGAAAAACAAAGAAA 592
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 612 AACCTGAAGAGAGAGCTAAAGAGAGCTAAAGAGAGAGCTAAAGAGCTAAAGAG 553
 QY 593 ATGGTGAGAAATTAATCTTGAATTCCTTCAACAGAGTACCTGAGCAAGTCAAAAAG 652
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 552 AACCTGAAGCAAAAGCTAAAGAGAGAGCTGAAGAGAGAGCTAAAGAGAGAGTAAAG 493
 QY 653 CCGCTTTCGCTTATGATGATGATGCTTCTGATGTGATTAAGCAAGCTGACACAG 712
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 492 AAAATAAGAGAGAGTTAAAGAGAGAGTTAAAGAGAGTTAAAGAGAGAGAG 433
 QY 713 CCTACAGTAGCCCAATTTACATTGTAAACTGCTACTAAGAAAAAGAAAAATCTTAATA 772
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 432 AAAATAAGAGAGAGAGTTAAAGAGAGAGTTAAAGAGAGAGTTAAAGAGAGAGAG 373
 QY 773 TTGATGACTCAAGAGAGAGTTAAAGAGAGTTAAAGAGAGAGTTAAAGAGAGAG 802
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 372 AAAATAAGAGAGAGTTAAAGAGAGAGTTAAAGAGAGAGTTAAAGAGAGAG 343
 RESULT 10
 A2530768 890 bp DNA linear GSS 03-NOV-2000
 LOCUS ENTB547F Entamoeba histolytica Sheared DNA Entamoeba histolytica
 DEFINITION genomic, genomic survey sequence.
 ACCESSION A2530768 GI:11084894
 VERSION A2530768
 KEYWORDS GSS.
 SOURCE Entamoeba histolytica
 ORGANISM Eukaryota; Entamoebidae; Entamoeba.
 REFERENCE 1 (bases 1 to 890)
 AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
 TITLE Determination of clone end sequences from Entamoeba histolytica
 HML:IMSS sheared DNA library
 JOURNAL Unpublished
 COMMENT Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b.loftus@tigr.org
 Clones are derived from the Entamoeba histolytica HML:IMSS sheared
 DNA library
 Seq primer: M13-Forward
 Class: shotgun
 High quality sequence start: 20
 High quality sequence stop: 822.
 Location/Qualifiers
 1..890
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HML:IMSS"
 /db_xref="taxon:5759"
 /note="Vector: PHOS1; Site 1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing. projects. In Genome
 Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Barrell, Oxford University Press, 1999)."
 BASE COUNT 375 a 61 c 225 g 229 t
 ORIGIN
 Query Match 6.7%; Score 64.8; DB 28; Length 890;
 Best Local Similarity 46.8%; Pred. No. 0.82;
 Matches 204; Conservative 0; Mismatches 232; Indels 0; Gaps 0;
 QY 218 AAGTTTGAAGAAATATGAGCTGATGATTAAGAGAGTTAAGAGAGTTAAGATCTATG 277
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 448 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 507
 QY 278 CCGAAG 337
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

D	508	AAGAAGTATGATGAAGATGATGAAGAATGATGATGAAGAAGATGATGAAGAATGATG	567
OY	338	CTCTTGAAACACGGTAAGCTCAATAATTGCATCAGACTAAATTAAGTTGAGTTGGCAGTTAAGA	397
D	568	ATGAAGATGATGATGAAGATGATGATGATGAAGACGATGATGAAGACGATGATGAAGACGATG	627
OY	398	AGGTACGAGAAAGCGATGATTCAGCATGATGAAGCGCTTAAGAAGCCCTTGATGATGATGAC	457
D	628	ATGAAGACGATGATGAAGACGATGATGAAGACGATGATGAAGACGATGATGAAGACGATG	687
OY	458	CAGATCTAACGGCTCAATCATCCGTTTAAATATGAAGATGAAGGCCAAAGAAAGTTTCG	517
D	688	ATGATGAAGATGAAGATGAAGACGATGATGATGAAGACGATGATGAAGACGATGATGATG	747
OY	518	AAAAAGCCAAAGCGAAGAGTCTGATTTTTGCTCAATTAGCCAAAGATTAATTCAACTGATG	577
D	748	AAGATGAAGACGATGATGATGAAGACGATGATGATGAAGACGATGATGAAGACGATG	807
OY	578	AAAAAACAAAAGAAATGCTGGAATAATTAACCTTGATTCGCTTCAACACAAGTACTG	637
D	808	ATGATGATGATGAAGAAAGAGAAAGACGCTTTAGAAAGACATCACAGAAACATTCG	867
OY	638	AGCAAGTCAAAAAAGC	653
D	868	GAGACGTTATGAAGAC	883

	RESULT	11
	BH149983	
LOCUS		
DEFINITION	BH149983	976 bp DNA linear GSS 27-AUG-2001
	ENTQD93JTF Entamoeba histolytica Sheared DNA	Entamoeba histolytica genomic, genomic survey sequence.
ACCESSION	BH149983	
VERSION	BH149983.1 GI:15311165	GSS.
KEYWORDS		
SOURCE		
ORGANISM	Entamoeba histolytica	
	Eukaryota; Eutrombicidaae; Entamoeba.	
REFERENCE	I (bases 1 to 976)	
AUTHORS	Lofthus,B., Wang,Z., Van Aken,S. and Fraser,C.	
TITLE	Determination of clone end sequences from Entamoeba histolytica HMI:IMSS sheared DNA library (2001)	
JOURNAL	Unpublished	
COMMENT	Contact: Brendan J Lofthus Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel.: 301 838 0208 Fax: 301 838 3543 Email: bjl@fusrstlgr.org	
	C clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library	
	Seq primer: M13-Forward	
	Class: shotgun	
	High quality sequence start: 38	
	High quality sequence stop: 740.	
FEATURES	Location/Qualifiers	
SOURCE	1..976	

```

FEATURES
    source
        Location/Qualifiers
            1..976
                /organism="Entamoeba histolytica"
                /mol_type="genomic DNA"
                /strain="HMI-1/MS"
                /db_xref="taxon:5759"
                /clone_lib="Entamoeba histolytica Sheared DNA"
                /note="Vector: pHS1.1; Site 1: Bst I; Constructed at The
                Institute for Genomic Research (TIGR), Rockville, MD.
                Genomic DNA isolated from broth cultures of E. histolytica
                using a method described by Clark and Diamond (Clark,
                C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
                method for isolate identification. Exp. Parasitol.
                77:450.). The DNA was mechanically sheared to give a
                tight size distribution (~2 kb). The v + i method used for
                the library construction is described in detail in Smith
                H.O. and Venter, J.C. (Making small insert libraries for

```

	Whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).							
BASE COUNT	415	a	63	c	280	g	218	t
ORIGIN								
Query Match				6.7%	Score	64.6;	DB	28;
Best Local Similarity				45.7%	Pred. No.	0.87;		
Matches	226;	Conservative		0;	Mismatches	269;	Indels	0;
							Gaps	0;

OY	194	TCCTGTTAAATATACCATCCAAAAGCTTTTGTAAAAACATATAGGCTACAGGCTGATG	253
Db	95	TATCTTAACTGCTTCATGCCGGAAGAGGTTGAACTCGATGATTTTGAATTTATA	154
OY	254	ATAAAGAGTTGATGATCTACTATTGCCGAGAAAAAACATATGCGGAAAACCTACAC	313
Db	155	ATGAAGATTTTGTATTTAAATGATGATGATGTATGTAAATTAGAAGACGATGAAAGAACGACG	214
OY	314	GTCGCTTGTCACAGCAGGATGACTGCTGTGAACACGTAAGCTCAAAATTTGTCACAACTA	373
Db	215	ATGATGATGAAATTGCAATTAGAACACGACGAAGAAGATGATGATGAAGAAGAACAGATG	274
OY	374	AATTAGTTGACTTGCGCAGTTAAGAAAGGTACGAAAGCTGAAATTGACAGATGAAAGCTTATA	433
Db	275	AAGAAGAAAGAAAGATGAAGAGMAAGATGAAAGATGAAGAAAGACATGATGAAGAG	334
OY	434	AGAAAGCCTTTGATGAGTACACCTCCAGATGTAAGCGCTCAATCATCCGCTTTAATAATG	493
Db	335	ACGATGATGAAGAAAGACGATGATGAAGAAGACGATGATGAAGAAAGACGATGATGAAGAG	394
OY	494	AAGATTAAGCCAAAGAAAGTTCTCGAAAAAGCCAAAGGCAGAAAGTGTGATTTTGTCTCAAT	553
Db	395	ACGATGATGAAGAAAGACGATGATGAAGAAGACGATGATGAAGAAAGACGATGATGAAGAG	454
OY	554	TAGCCAAAGATTAATTCACATGATGTAAGAAAACAAAGAAATAGTGTGGAATTAATCTTTG	613
Db	455	ACGATGATGAAGAAAGACGATGATGTAAGAAAGCAGATGATGAAGAAAGATGATGAAGAGT	514
OY	614	ATTCTGCTTCAACAGAAAGTACCTAGCAAGTCAAAAAAGCCGCTTTCGCTTTAGATGTGG	673
Db	515	ATGATGATGAAGATGATGAATTTGTAAATTAGAAGACGATGATGAAGATGATGATGAAGAG	574
OY	674	ATGGATGTTCTGATG	688
Db	575	ATGATGATGAAGATG	589

RESULT 12
 A2669474
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

A2669474
 ENT1188TR Entamoeba histolytica Sheared DNA
 genomic, genomic survey sequence.
 A2669474
 A2669474.1 GI:11806620
 GSS.
 Entamoeba histolytica
 Entamoeba histolytica
 Eukaryota: Entamebidae; Entamoeba.
 1 (bases 1 to 880)
 Loftus,B., Van Aken,S. and Fraser,C.
 Determination of clone end sequences from Entamoeba histolytica
 HM1:IMSS sheared DNA library
 Unpublished
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjoftus@tigr.org
 Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
 DNA library
 Seq primer: M13-Reverse


```
Db      464 AGATGATGACGTAGATTAATGATGACGAAAGATGACGCTAGAT--GATGATGACGAGA 407
QY      393 TAAGAAGGTAGCAGAAAGCTGAATTGACATGAGAGCCTTAAGAAAGCCTTGATGAGTA 452
Db      406 TGATGACGAAAGATGATGATGACGAAATGATGACGATGATGATGACGAAAGATGA 347
QY      453 CACTCCAGATGTAAAGGCTCAATCATCCGCTTATTAATGAAGATAAGGCCAAAGAGT 512
Db      346 CGAAGATGATGAAGAAGATGATGACCTAGATGATGACGATGATGATGACGAAAGATGA 287
QY      513 TCTCGAAAAAGCCAGAGGCTGCTGATTTGCTCAATTAGCCAAAGATAATTCAAC 572
Db      286 TGACGAAAGATGATGACGTAGATGATGACGAAATGATGACGAAAGATGATGACGAAAG 227
QY      573 TGATGAAAAACAAAAAATGCTGAGAAATTAACCTTGATTCCTTCAACAGAGT 632
Db      226 TGATGACGAAAGATGACGAAAGATGATGACGACGATGATGATGATGATGATGATGATGA 167
QY      633 ACCTGAGCAAGTCMAAAAAAGCCCTTTCGCTTATGATGATGATGATGATGATGATGAT 692
Db      166 TGACAAAGATGACGAAAGATGCTGACGAAGATGATGATGATGATGATGATGATGATGAT 107
QY      693 T 693
Db      106 T 106
```

Search completed: September 4, 2003, 21:02:05
Job time : 2074 secs

BEST AVAILABLE COPY

THIS PAGE BLANK (USPTO)

Fri Sep 5 09:24:49 2003

us-10-049-473a-1.inpm

Graser
10/049473 Page 1
Seq. 105 1st 2nd
Interf

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2003, 19:26:32 ; Search time 2926 Seconds

(without alignments)
10818.092 Million cell updates/sec

Title: US-10-049-473a-1

Perfect score: 966

Sequence: 1 agtaacattatctcaag.....gcagttagtactcaacgaa 966

Scoring table: IDENTITY_NTC
Gapop 10.0 , Gapext 1.0

Searched: 32997241 seqs, 16383922548 residues 65994482

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/1/pna/US100A.COMB.seq.*
2: /cgn2_6/ptodata/1/pna/US100B.COMB.seq.*
3: /cgn2_6/ptodata/1/pna/US100C.COMB.seq.*
4: /cgn2_6/ptodata/1/pna/US100D.COMB.seq.*
5: /cgn2_6/ptodata/1/pna/US100E.COMB.seq.*
6: /cgn2_6/ptodata/1/pna/US100F.COMB.seq.*
7: /cgn2_6/ptodata/1/pna/US100G.COMB.seq.*
8: /cgn2_6/ptodata/1/pna/US100H.COMB.seq.*
9: /cgn2_6/ptodata/1/pna/US100I.COMB.seq.*
10: /cgn2_6/ptodata/1/pna/US100J.COMB.seq.*
11: /cgn2_6/ptodata/1/pna/US100K.COMB.seq.*
12: /cgn2_6/ptodata/1/pna/US100L.COMB.seq.*
13: /cgn2_6/ptodata/1/pna/US100M.COMB.seq.*
14: /cgn2_6/ptodata/1/pna/US100N.COMB.seq.*
15: /cgn2_6/ptodata/1/pna/US100O.COMB.seq.*
16: /cgn2_6/ptodata/1/pna/US100P.COMB.seq.*
17: /cgn2_6/ptodata/1/pna/US100Q.COMB.seq.*
18: /cgn2_6/ptodata/1/pna/US100R.COMB.seq.*
19: /cgn2_6/ptodata/1/pna/US100S.COMB.seq.*
20: /cgn2_6/ptodata/1/pna/US100T.COMB.seq.*
21: /cgn2_6/ptodata/1/pna/US100U.COMB.seq.*
22: /cgn2_6/ptodata/1/pna/US100V.COMB.seq.*
23: /cgn2_6/ptodata/1/pna/US100W.COMB.seq.*
24: /cgn2_6/ptodata/1/pna/US100X.COMB.seq.*
25: /cgn2_6/ptodata/1/pna/US100Y.COMB.seq.*
26: /cgn2_6/ptodata/1/pna/US100Z.COMB.seq.*
27: /cgn2_6/ptodata/1/pna/US100A.COMB.seq.*
28: /cgn2_6/ptodata/1/pna/US100B.COMB.seq.*
29: /cgn2_6/ptodata/1/pna/US100C.COMB.seq.*
30: /cgn2_6/ptodata/1/pna/US100D.COMB.seq.*
31: /cgn2_6/ptodata/1/pna/US100E.COMB.seq.*
32: /cgn2_6/ptodata/1/pna/US100F.COMB.seq.*
33: /cgn2_6/ptodata/1/pna/US100G.COMB.seq.*
34: /cgn2_6/ptodata/1/pna/US100H.COMB.seq.*
35: /cgn2_6/ptodata/1/pna/US100I.COMB.seq.*
36: /cgn2_6/ptodata/1/pna/US100J.COMB.seq.*
37: /cgn2_6/ptodata/1/pna/US100K.COMB.seq.*
38: /cgn2_6/ptodata/1/pna/US100L.COMB.seq.*
39: /cgn2_6/ptodata/1/pna/US100M.COMB.seq.*
40: /cgn2_6/ptodata/1/pna/US100N.COMB.seq.*
41: /cgn2_6/ptodata/1/pna/US100O.COMB.seq.*
42: /cgn2_6/ptodata/1/pna/US100P.COMB.seq.*
43: /cgn2_6/ptodata/1/pna/US100Q.COMB.seq.*

44: /cgn2_6/ptodata/1/pna/US100R.COMB.seq.*
45: /cgn2_6/ptodata/1/pna/US100S.COMB.seq.*
46: /cgn2_6/ptodata/1/pna/US100T.COMB.seq.*
47: /cgn2_6/ptodata/1/pna/US100U.COMB.seq.*
48: /cgn2_6/ptodata/1/pna/US100V.COMB.seq.*
49: /cgn2_6/ptodata/1/pna/US100W.COMB.seq.*
50: /cgn2_6/ptodata/1/pna/US100X.COMB.seq.*
51: /cgn2_6/ptodata/1/pna/US100Y.COMB.seq.*
52: /cgn2_6/ptodata/1/pna/US100Z.COMB.seq.*
53: /cgn2_6/ptodata/1/pna/US100A.COMB.seq.*
54: /cgn2_6/ptodata/1/pna/US100B.COMB.seq.*
55: /cgn2_6/ptodata/1/pna/US100C.COMB.seq.*
56: /cgn2_6/ptodata/1/pna/US100D.COMB.seq.*
57: /cgn2_6/ptodata/1/pna/US100E.COMB.seq.*
58: /cgn2_6/ptodata/1/pna/US100F.COMB.seq.*
59: /cgn2_6/ptodata/1/pna/US100G.COMB.seq.*
60: /cgn2_6/ptodata/1/pna/US100H.COMB.seq.*
61: /cgn2_6/ptodata/1/pna/US100I.COMB.seq.*
62: /cgn2_6/ptodata/1/pna/US100J.COMB.seq.*
63: /cgn2_6/ptodata/1/pna/US100K.COMB.seq.*
64: /cgn2_6/ptodata/1/pna/US100L.COMB.seq.*
65: /cgn2_6/ptodata/1/pna/US100M.COMB.seq.*
66: /cgn2_6/ptodata/1/pna/US100N.COMB.seq.*
67: /cgn2_6/ptodata/1/pna/US100O.COMB.seq.*
68: /cgn2_6/ptodata/1/pna/US100P.COMB.seq.*
69: /cgn2_6/ptodata/1/pna/US100Q.COMB.seq.*
70: /cgn2_6/ptodata/1/pna/US100R.COMB.seq.*
71: /cgn2_6/ptodata/1/pna/US100S.COMB.seq.*
72: /cgn2_6/ptodata/1/pna/US100T.COMB.seq.*
73: /cgn2_6/ptodata/1/pna/US100U.COMB.seq.*
74: /cgn2_6/ptodata/1/pna/US100V.COMB.seq.*
75: /cgn2_6/ptodata/1/pna/US100W.COMB.seq.*
76: /cgn2_6/ptodata/1/pna/US100X.COMB.seq.*
77: /cgn2_6/ptodata/1/pna/US100Y.COMB.seq.*
78: /cgn2_6/ptodata/1/pna/US100Z.COMB.seq.*
79: /cgn2_6/ptodata/1/pna/US100A.COMB.seq.*
80: /cgn2_6/ptodata/1/pna/US100B.COMB.seq.*
81: /cgn2_6/ptodata/1/pna/US100C.COMB.seq.*
82: /cgn2_6/ptodata/1/pna/US100D.COMB.seq.*
83: /cgn2_6/ptodata/1/pna/US100E.COMB.seq.*
84: /cgn2_6/ptodata/1/pna/US100F.COMB.seq.*
85: /cgn2_6/ptodata/1/pna/US100G.COMB.seq.*
86: /cgn2_6/ptodata/1/pna/US100H.COMB.seq.*
87: /cgn2_6/ptodata/1/pna/US100I.COMB.seq.*
88: /cgn2_6/ptodata/1/pna/US100J.COMB.seq.*
89: /cgn2_6/ptodata/1/pna/US100K.COMB.seq.*
90: /cgn2_6/ptodata/1/pna/US100L.COMB.seq.*
91: /cgn2_6/ptodata/1/pna/US100M.COMB.seq.*
92: /cgn2_6/ptodata/1/pna/US100N.COMB.seq.*
93: /cgn2_6/ptodata/1/pna/US100O.COMB.seq.*
94: /cgn2_6/ptodata/1/pna/US100P.COMB.seq.*
95: /cgn2_6/ptodata/1/pna/US100Q.COMB.seq.*
96: /cgn2_6/ptodata/1/pna/US100R.COMB.seq.*
97: /cgn2_6/ptodata/1/pna/US100S.COMB.seq.*
98: /cgn2_6/ptodata/1/pna/US100T.COMB.seq.*
99: /cgn2_6/ptodata/1/pna/US100U.COMB.seq.*
100: /cgn2_6/ptodata/1/pna/US100V.COMB.seq.*
101: /cgn2_6/ptodata/1/pna/US100W.COMB.seq.*
102: /cgn2_6/ptodata/1/pna/US100X.COMB.seq.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	966	100.0	966	US-10-049-473a-1	Sequence 1, App1
2	966	100.0	11309	US-10-158-844-108	Sequence 108, App
3	965.2	99.9	11167	US-60-029-960-148	Sequence 148, App
4	961.2	99.5	12356	US-60-068-175-584	Sequence 584, App

```
5 939 97.2 942 33 US-09-760-787-179 Sequence 179, App
6 934.2 96.9 942 25 US-09-583-110-2539 Sequence 2539, Ap
7 868.2 89.9 888 16 US-09-107-433-1524 Sequence 1524, Ap
8 807.8 83.6 11591 60 US-60-061-998-559 Sequence 559, App
9 552.4 57.2 640 14 US-08-961-083-33 Sequence 33, App
10 552.4 57.2 640 33 US-09-765-271-33 Sequence 33, App
11 552.4 57.2 640 33 US-09-765-272-33 Sequence 33, App
12 552.4 57.2 640 33 US-09-765-272A-33 Sequence 33, App
13 265.6 27.5 451 13 PCT-US97-07950-219 Sequence 219, App
14 265.6 27.5 791 1 PCT-US97-07950-219 Sequence 219, App
15 265.6 27.5 791 2 PCT-US97-07950-219 Sequence 219, App
16 216.2 22.4 1110 48 US-10-206-576-181 Sequence 181, App
17 215.6 22.3 1107 16 US-09-134-000C-879 Sequence 879, App
18 215.6 22.3 1107 16 US-10-434-665-879 Sequence 879, App
19 215.6 22.3 1107 53 US-10-434-665-879 Sequence 879, App
20 215.6 22.3 12860 15 US-09-070-927-144 Sequence 144, App
21 215.6 22.3 12860 15 US-09-070-927A-144 Sequence 144, App
22 211.6 21.9 2021 60 US-60-068-217-711 Sequence 711, App
23 209.8 21.7 1035 52 US-10-417-884-2701 Sequence 2701, App
24 208.8 21.6 2102 59 US-60-050-444-451 Sequence 451, App
25 208.8 21.6 2102 59 US-60-050-444-451 Sequence 451, App
26 207.2 21.4 7169 60 US-60-050-444-451 Sequence 451, App
27 207.2 21.4 7227 59 US-60-050-444-451 Sequence 451, App
28 206.4 21.4 912 52 US-10-417-884-2459 Sequence 2459, App
29 203.2 21.0 970 52 US-10-417-884-2459 Sequence 2459, App
30 151.2 15.7 909 52 US-10-417-884-1082 Sequence 1082, App
31 149.6 15.5 15081 60 US-60-068-186-732 Sequence 732, App
32 149.6 15.5 15497 59 US-60-046-553-782 Sequence 782, App
33 142.4 14.7 1599 58 US-60-046-553-782 Sequence 782, App
34 142.4 14.7 2231 58 US-60-045-549-843 Sequence 843, App
35 79.2 8.2 17475 29 US-09-663-779-1357 Sequence 1357, App
36 76.6 7.9 43980 51 US-10-398-221-5 Sequence 5, App
37 76.6 7.9 3011208 51 US-10-398-221-2058 Sequence 2058, App
38 73.4 7.6 340 58 US-60-045-549-1441 Sequence 1441, App
39 73.4 7.6 340 58 US-60-045-549-1441 Sequence 1441, App
40 71.8 7.4 1720 51 US-10-398-221-3334 Sequence 3334, App
41 59.8 6.2 7218 9 US-08-466-194-14 Sequence 14, App
42 59.2 6.1 1163 51 US-10-398-221-1250 Sequence 1250, App
43 53.8 5.6 2205 49 US-10-266-090-32333 Sequence 32333, A
44 53.8 5.6 2135 49 US-10-282-122A-16758 Sequence 16758, A
45 53.4 5.5 1206 49 US-10-266-090-24388 Sequence 24388, A
```

ALIGNMENTS

```
RESULT 1
US-10-049-473a-1
: Sequence 1, Application US/10049473a
: GENERAL INFORMATION:
: APPLICANT: de Groot, Ronald
: TITLE OF INVENTION: Pneumococcal Vaccines
: FILE REFERENCE: Docket 294-120 PCT/US
: CURRENT APPLICATION NUMBER: US/10/049, 473a
: PCT FILING DATE: 2002-02-12
: PRIOR APPLICATION NUMBER: PCT/NL00/00569
: PRIOR FILING DATE: 2000-08-14
: PRIOR APPLICATION NUMBER: EP 99202640.1
: NUMBER OF SEQ ID NOS: 5
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 1
: LENGTH: 966
: TYPE: DNA
: ORGANISM: Streptococcus pneumoniae
: US-10-049-473a-1
```

```
Query Match 100.0%; Score 966; DB 44; Length 966;
Best Local Similarity 100.0%; Pred. No. 8.9e-186;
Matches 966; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CY 1 AGTAACCTTATCTCAAGAGTAGACATGAGAAAAAATTTATGCGAGGTGCATCACA 60
```

```
Db
1 AGTAACCTTATCTCAAGAGTAGACATGAGAAAAAATTTATGCGAGGTGCATCACA 60
61 CTATTATCAGTAGCACTTATAGCAGCTTGTTCGAAAGGTCAGAAAGGTGCAGACCTATC 120
61 CTATTATCAGTAGCACTTATAGCAGCTTGTTCGAAAGGTCAGAAAGGTGCAGACCTATC 120
121 AGCATGAAAGGAGTGCATTTACAGAACATCAATTTATGAGCAAGTGAAGCAACCT 180
121 AGCATGAAAGGAGTGCATTTACAGAACATCAATTTATGAGCAAGTGAAGCAACCT 180
181 TCAGCCCAACAGCTTGTATTAATATGACCATCCAAAGTTTGTGAAGCAATATGCG 240
181 TCAGCCCAACAGCTTGTATTAATATGACCATCCAAAGTTTGTGAAGCAATATGCG 240
241 TCAGAGCTTGTATGATTAAGAGCTTGTATGATGATGATGATGATGATGATGATG 300
241 TCAGAGCTTGTATGATTAAGAGCTTGTATGATGATGATGATGATGATGATGATG 300
301 GAAAGCTACCAAGCTGCTGTGTCACAGAGGTATGCTGTGTAACAGCTTAAGCTCA 360
301 GAAAGCTACCAAGCTGCTGTGTCACAGAGGTATGCTGTGTAACAGCTTAAGCTCA 360
361 ATTGCTACAGTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 420
361 ATTGCTACAGTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 420
421 GATGAGGCTTATGAGAAAGCTTGTGATGATGATGATGATGATGATGATGATGATG 480
421 GATGAGGCTTATGAGAAAGCTTGTGATGATGATGATGATGATGATGATGATGATG 480
481 GCTTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
481 GCTTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
541 GATTTGCTCAATTAAGCAAGTAATTAATTAATTAATTAATTAATTAATTAATTAAT 600
541 GATTTGCTCAATTAAGCAAGTAATTAATTAATTAATTAATTAATTAATTAATTAAT 600
601 GAAATTAACCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
601 GAAATTAACCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
661 GCTTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720
661 GCTTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720
721 AGCAATATTACATTTATTAACCTCACTAGAAAGCAAGTAATTAATTAATTAATTAAT 780
721 AGCAATATTACATTTATTAACCTCACTAGAAAGCAAGTAATTAATTAATTAATTAAT 780
781 TACAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 840
781 TACAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 840
841 CAAGCATTAATGAGAAAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 900
841 CAAGCATTAATGAGAAAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 900
901 CAAATATCTTATACCAATATATATGATGATGATGATGATGATGATGATGATGATGAT 960
901 CAAATATCTTATACCAATATATATGATGATGATGATGATGATGATGATGATGATGAT 960
961 AACGAA 966
961 AACGAA 966
Db
```

```
RESULT 2
US-10-158-844-108
: Sequence 108, Application US/10158844
: GENERAL INFORMATION:
: APPLICANT: Kunsch et al.
```



```

? TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
? NUMBER OF SEQUENCES: 391
? CORRESPONDENCE ADDRESS:
? ADDRESSER: Human Genome Sciences, Inc.
? STREET: 9410 Key West Avenue
? CITY: Rockville
? STATE: Maryland
? COUNTRY: USA
? ZIP: 20850
? COMPUTER READABLE FORM:
? MEDIUM TYPE: CD-R
? COMPUTER: Dell Latitude Pentium 3
? OPERATING SYSTEM: Windows 98
? SOFTWARE: ASCII Text
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/10/158,844
? FILING DATE: 03-Jun-2002
? CLASSIFICATION: <unknown>
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/961,527
? FILING DATE: 1997-10-30
? APPLICATION NUMBER: US 60/029,960
? FILING DATE: 1996-10-31
? ATTORNEY/AGENT INFORMATION:
? NAME: Hyman, Mark J.
? REGISTRATION NUMBER: 46,789
? REFERENCE/DOCKET NUMBER: PB340P1D1
? INFORMATION FOR SEQ ID NO: 108:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 11309 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? SEQUENCE DESCRIPTION: SEQ ID NO: 108:
US-10-158-844-108

Query Match      100.0%; Score 966; DB 47; Length 11309;
Best Local Similarity 100.0%; Pred. No. 1.6e-187;
Matches 966; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTAACACTTATCTCAAGAGATGACATCAAGAAAAAATATTTGGCAGGTGCCATCACA 60
DB 7632 AGTAACACTTATCTCAAGAGATGACATCAAGAAAAAATATTTGGCAGGTGCCATCACA 7691
QY 61 CTAATATCAAGTACCACTTTAGCAGCTTGTTCGAAAGGTCAGAGAAGTGCAGACTTATC 120
DB 7692 CTAATATCAAGTACCACTTTAGCAGCTTGTTCGAAAGGTCAGAGAAGTGCAGACTTATC 7751
QY 121 AGCATGAAAGGGGATGTCATTACAGAACATCAATTTTATGAGCAAGTGAAGAACCAACCT 180
DB 7752 AGCATGAAAGGGGATGTCATTACAGAACATCAATTTTATGAGCAAGTGAAGAACCAACCT 7811
QY 181 TCAGCCCAACAATCTTGTAAATATGACCATCCAAAAAGTTTGAAGAAACAATATGGC 240
DB 7812 TCAGCCCAACAATCTTGTAAATATGACCATCCAAAAAGTTTGAAGAAACAATATGGC 7871
QY 241 TCAGACTTGATGATTAAGAGTGTGATGATCTATTGCCGAAGAAAAAACAATATGGC 300
DB 7872 TCAGACTTGATGATTAAGAGTGTGATGATCTATTGCCGAAGAAAAAACAATATGGC 7931
QY 301 GAAAGTACCAACGCTGCTTGTCAACAAGCATGATCTTTGAAACAGCTAAAGCTCAA 360
DB 7932 GAAAGTACCAACGCTGCTTGTCAACAAGCATGATCTTTGAAACAGCTAAAGCTCAA 7991
QY 361 ATTCGTACAAGTAATAGTTAGTTGGCAGTTAAGAAGTACAGAAAGCTGAATTGACA 420
DB 7992 ATTCGTACAAGTAATAGTTAGTTGGCAGTTAAGAAGTACAGAAAGCTGAATTGACA 8051
QY 421 GATGAAGCCATTAAGAAAGCCTTGTGATGATGATCACTCCAGATGTAAGCGCTCAATCATC 480
DB 8052 GATGAAGCCATTAAGAAAGCCTTGTGATGATGATCACTCCAGATGTAAGCGCTCAATCATC 8111
QY 481 CGCTTATATATGATGATTAAGGCCAAGAAAGTTCTCGAAAAAGCCAAAGGAGAGTGTCT 540

```

```

DB 8112 CGCTTATATGATGATTAAGGCCAAGAAAGTTCTCGAAAAAGCCAAAGCAGAGAGTGTCT 8171
QY 541 GATTTCGTCATATTACCAAGATATTTCACTGATGAAAAAACAAGAAATGCTGCA 600
DB 8172 GATTTCGTCATATTACCAAGATATTTCACTGATGAAAAAACAAGAAATGCTGCA 8231
QY 601 GAAATACCTTTGATCTGCTTCACAGAAAGTACGTAGCAAGTCAAAAAAGCGCTTTC 660
DB 8232 GAAATACCTTTGATCTGCTTCACAGAAAGTACGTAGCAAGTCAAAAAAGCGCTTTC 8291
QY 661 GCTTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
DB 8292 GCTTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8351
QY 721 AGCCATATTTACCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
DB 8352 AGCCATATTTACCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8411
QY 781 TACAAGAAAAAATTAAGAAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 840
DB 8412 TACAAGAAAAAATTAAGAAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 8471
QY 841 CAAAGCATTTATCGGAAAGATTTGCAAGCAGCCAAATATCAAGGTTAAGACCAAGCCTTC 900
DB 8472 CAAAGCATTTATCGGAAAGATTTGCAAGCAGCCAAATATCAAGGTTAAGACCAAGCCTTC 8531
QY 901 CAAATATCTTTTCCCAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
DB 8532 CAAATATCTTTTCCCAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8591
QY 961 AACGAA 966
DB 8592 AACGAA 8597

RESULT 3
US-60-029-960-148/c
? Sequence 148, Application US/60029960
? GENERAL INFORMATION:
? APPLICANT: Charles Kunsch
? TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
? NUMBER OF SEQUENCES: 1649
? CORRESPONDENCE ADDRESS:
? ADDRESSER: Human Genome Sciences, Inc.
? STREET: 9410 Key West Avenue
? CITY: Rockville
? STATE: Maryland
? COUNTRY: USA
? ZIP: 20850
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette, 3.50 Inch, 1.4MB storage
? COMPUTER: HP Vectra 486/33
? OPERATING SYSTEM: MSDOS version 6.2
? SOFTWARE: ASCII Text
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/60/029,960
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: Brookes, A. Anders
? REGISTRATION NUMBER: 36,373
? REFERENCE/DOCKET NUMBER: PB340PP
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (301) 309-8504
? TELEFAX: (301) 309-8512
? INFORMATION FOR SEQ ID NO: 148:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 11167 base pairs
? TYPE: nucleic acid

```

STRANDEDNESS: double
TOPOLOGY: Linear
US-60-029-960-148

Query Match 99.9%; Score 965.2; DB 56; Length 11167;
Best Local Similarity 99.8%; Pred. No. 2.3e-187;
Matches 964; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 AGTAACACTTATCTCAAAAGAGTAGACATGAGAAAAAATTTATGGCAGGTGCCATCACA 60
DB 3693 AGTAACACTTATCTCAAAAGAGTAGACATGAGAAAAAATTTATGGCAGGTGCCATCACA 3634
QY 61 CTATTATCACTAGCAACTTTAGCAGCTTGTTCGAAAGGTCAGAGAGTGCAGACCTTATC 120
DB 3633 CTATTATCACTAGCAACTTTAGCAGCTTGTTCGAAAGGTCAGAGAGTGCAGACCTTATC 3574
QY 121 AGCATGAAGGGAGTGCATTAAGCAATCAATTTATGAGCAAGTGAAGAACCTT 180
DB 3573 AGCATGAAGGGAGTGCATTAAGCAATCAATTTATGAGCAAGTGAAGAACCTT 3514
QY 181 TCAGCCCAACAGTCTTGTAAATATGACCATCCAAAAGTTTGTGAAAAACAATATGCG 240
DB 3513 TCAGCCCAACAGTCTTGTAAATATGACCATCCAAAAGTTTGTGAAAAACAATATGCG 3454
QY 241 TCAGAGCTTATGATTAAGAGGTTGATGATCTATTTGCCAGAAAAACAATATGCG 300
DB 3453 TCAGAGCTTATGATTAAGAGGTTGATGATCTATTTGCCAGAAAAACAATATGCG 3394
QY 301 GAAACACTACCAAGTGTCTGTGTCACAAGAGTATGACTCTTGAACACGTAAGAGCTCAA 360
DB 3393 GAAACACTACCAAGTGTCTGTGTCACAAGAGTATGACTCTTGAACACGTAAGAGCTCAA 3334
QY 361 ATTGCTACAAAGTAATTAAGTTGAGTTGGCAGTTAAGAGGTAGCAGAGGTGAATTGACA 420
DB 3333 ATTGCTACAAAGTAATTAAGTTGAGTTGGCAGTTAAGAGGTAGCAGAGGTGAATTGACA 3274
QY 421 GATGAAGCCATATAGAAAGCCTTGTGATGATACCTCAGATGAACGGGTCAAAATATGAC 480
DB 3273 GATGAAGCCATATAGAAAGCCTTGTGATGATACCTCAGATGAACGGGTCAAAATATGAC 3214
QY 481 CGTCTTAATATAGTAAGTAAGGCCAAAGAAAGTTCTCGAAAAAGCCCAAGAGAGGTGCT 540
DB 3213 CGTCTTAATATAGTAAGTAAGGCCAAAGAAAGTTCTCGAAAAAGCCCAAGAGAGGTGCT 3154
QY 541 GATTTGCTCAATTAAGCAAGATTAATCAACTGATGAAAAAACAAGAAATGGTGGCA 600
DB 3153 GATTTGCTCAATTAAGCAAGATTAATCAACTGATGAAAAAACAAGAAATGGTGGCA 3094
QY 601 GAAATTACCTTGTATCTGCTTCAACAGAGTACTGAGCAAGTCAAAAAAGCCGCTTTC 660
DB 3093 GAAATTACCTTGTATCTGCTTCAACAGAGTACTGAGCAAGTCAAAAAAGCCGCTTTC 3034
QY 661 GCTTTAGAGTGCAGTGTGTTCTGATGATTAAGCAACTGGCACACCAAGCCCTACGT 720
DB 3033 GCTTTAGAGTGCAGTGTGTTCTGATGATTAAGCAACTGGCACACCAAGCCCTACGT 2974
QY 721 AGCCAAATTTACATTTGTAAGAACCTCACTAAGAAAAAGAAATCATCTAATATTGATGAC 780
DB 2973 AGCCAAATTTACATTTGTAAGAACCTCACTAAGAAAAAGAAATCATCTAATATTGATGAC 2914
QY 781 TACAAGAAAAAATTAATAAACTGTTATCTGACTCAAAAAACAATGATTAACATTTGTT 840
DB 2913 TACAAGAAAAAATTAATAAACTGTTATCTGACTCAAAAAACAATGATTAACATTTGTT 2854
QY 841 CAAAGCATTTATCGAAAAAGATTCAGACGACCAATATCAAGGTTAAGGACCAAGCCCTTC 900
DB 2853 CAAAGCATTTATCGAAAAAGATTCAGACGACCAATATCAAGGTTAAGGACCAAGCCCTTC 2794
QY 901 CAAAATATCTTTCCCAATATATCGGTGGTGGAGATTCAAGCTCAAGCAGTATGATCA 960
DB 2793 CAAAATATCTTTCCCAATATATCGGTGGTGGAGATTCAAGCTCAAGCAGTATGATCA 2734
QY 961 AACGAA 966

```

DB 2733 AACGAA 2728

RESULT 4

US-60-068-175-584

Sequence 584, Application US/60068175

GENERAL INFORMATION:

APPLICANT: Lagace, Robert E.

APPLICANT: Corley, Neil C.

APPLICANT: Russo, Frank D.

APPLICANT: Hann, Amy L.

APPLICANT: Finney, Gregory L.

TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF STREPTOCOCCUS PNEUMONIAE

TITLE OF INVENTION: GENOME, FRAGMENTS THEREOF, AND USES THEREOF

NUMBER OF SEQUENCES: 1175

CORRESPONDENCE ADDRESS:

ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

STREET: 3174 PORTER DRIVE

CITY: PALO ALTO

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/60/068,175

FILING DATE: HEREWITH

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: CERRONE, MICHAEL C.

REGISTRATION NUMBER: 39,132

REFERENCE/DOCKET NUMBER: PM-0009-2 P

TELECOMMUNICATION INFORMATION:

TELEPHONE: (650) 855-0555

TELEFAX: (650) 845-416

INFORMATION FOR SEQ ID NO: 584:

SEQUENCE CHARACTERISTICS:

LENGTH: 12356 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: genomic DNA

IMMEDIATE SOURCE:

CLONE: SPN2c586

US-60-068-175-584

Query Match

Best Local Similarity 99.5%; Score 961.2; DB 60; Length 12356;

Matches 963; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 1 AGTAACACTTATCTCAAAAGAGTAGACATGAGAAAAAATTTATGGCAGGTGCCATCACA 60
DB 8278 AGTAACACTTATCTCAAAAGAGTAGACATGAGAAAAAATTTATGGCAGGTGCCATCACA 8337
QY 61 CTATTATCACTAGCAACTTTAGCAGCTTGTTCGAAAGGTCAGAGAGTGCAGACCTTATC 120
DB 8338 CTATTATCACTAGCAACTTTAGCAGCTTGTTCGAAAGGTCAGAGAGTGCAGACCTTATC 8397
QY 121 AGCATGAAGGGAGTGCATTAAGCAATCAATTTATGAGCAAGTGAAGAACCAACCTT 180
DB 8398 AGCATGAAGGGAGTGCATTAAGCAATCAATTTATGAGCAAGTGAAGAACCAACCTT 8457
QY 181 TCAGCCCAACAGTCTTGTAAATATGACCATCCAAAAGTTTGTGAAAAACAATATGCG 240
DB 8458 TCAGCCCAACAGTCTTGTAAATATGACCATCCAAAAGTTTGTGAAAAACAATATGCG 8517
QY 241 TCAGAGCTTATGATTAAGAGGTTGATGATCTATTTGCCAGAAAAACAATATGCG 300

```

```

RESULT 5
US-09-769-787-179
; Sequence 179: Application US/09769787
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PMC/P211290
; CURRENT APPLICATION NUMBER: US/09/769,787
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 179
; LENGTH: 942
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; US-09-769-787-179

```

QY	28	ATGAGAAAAAATTTATTTGGCAGAGTGCCTATCAGACTATATATGATGCAACTTTACAGCT	87
Db	1	ATGAGAAAAAATTTATTTGGCAGAGTGCCTATCAGACTATATATGATGCAACTTTACAGCT	60
QY	88	TGTTTGAAGGGGTGAGAGGTGCAGACCTTATCAGCATGAAAGGGATGTCTATACAGAA	147
Db	61	TGTTTGAAGGGGTGAGAGGTGCAGACCTTATCAGCATGAAAGGGATGTCTATACAGAA	120
QY	148	CATCAATTTTATGAGCAAGTAAAAAGCAACCCCTTCAGGCCAAACAGTCTTGTAAATATG	207
Db	121	CATCAATTTTATGAGCAAGTAAAAAGCAACCCCTTTCAGGCCAAACAGTCTTGTAAATATG	180
QY	208	ACCATCCAAAAAGTTTTGGAAAAAACAAATNTGGCTCAGAGCTTGATGATTAAGAGGGTTGAT	267
Db	181	ACCATCCAAAAAGTTTTGGAAAAAACAAATNTGGCTCAGAGCTTGATGATTAAGAGGGTTGAT	240
QY	268	GATACTATTTGCCGAGAAAAAAAACAAATATGGCCAAAACATACCAACGTGTCTTGCACAA	327
Db	241	GATACTATTTGCCGAGAAAAAAAACAAATNTGGCCAAAACATACCAACGTGTCTTGCACAA	300
QY	328	GCAGGTATGACTCTTGAAACAGCTTAACCTCAAAATTCGTACAGTAATTAAGTATGATGG	387
Db	301	GCAGGTATGACTCTTGAAACAGCTTAACCTCAAAATTCGTACAGTAATTAAGTATGATGG	360
QY	388	GCAGTTAAGAGAGTACGAGAAAGCTGAAATTGACAGATGGAAGCCTATTAAGAAAGCCTTGTAT	447
Db	361	GCAGTTAAGAGAGTACGAGAAAGCTGAAATTGACAGATGGAAGCCTATTAAGAAAGCCTTGTAT	420
QY	448	GAGTACACTCCAGATGTAAAGGCTCAAAATCATCGTCTTAAATATGAAGATTAAGGCCAAA	507
Db	421	GAGTACACTCCAGATGTAAAGGCTCAAAATCATCGTCTTAAATATGAAGATTAAGGCCAAA	480
QY	508	GAACTTTCGAAAAAGCCCAAGGCAAGAGTGTGATTTTGCATATTTAGGCCAAAGATAT	567
Db	481	GAACTTTCGAAAAAGCCCAAGGCAAGAGTGTGATTTTGCATATTTAGGCCAAAGATAT	540
QY	568	TCAACTGATGAAAAAACAAAAAGAAATGTTGAGAGAAATTAACCTTTGATTCGCTTCAACA	627
Db	541	TCAACTGATGAAAAAACAAAAAGAAATGTTGAGAGAAATTAACCTTTGATTCGCTTCAACA	600
QY	628	GAACTACTGAGCAAGTCAAAAAAGCCGCTTTCGCTTTAGATGTGAGATGTGTTCTGAT	687
Db	601	GAACTACTGAGCAAGTCAAAAAAGCCGCTTTCGCTTTAGATGTGAGATGTGTTCTGAT	660
QY	688	GTGATTCACGAACGTGGCACAACGACCCATACAGTACCAATATTAATGTAAACCTCACT	747
Db	661	GTGATTCACGAACGTGGCACAACGACCCATACAGTACCAATATTAATGTAAACCTCACT	720
QY	748	AAGAAAAACGAAAAAATCATCTAAATTTATGATGACTCAAAAGAAAAATTTAAAAACCTTATC	807
Db	721	AAGAAAAACGAAAAAATCATCTAAATTTATGATGACTCAAAAGAAAAATTTAAAAACCTTATC	780
QY	808	TTGACTCAAAAACAAAATGATTTCAACATTTGTTCAACAGCATTTATCGAAAAAGATTTGCA	867
Db	781	TTGACTCAAAAACAAAATGATTTCAACATTTGTTCAACAGCATTTATCGAAAAAGATTTGCA	840
QY	868	GCAGCAATATCAAGGTTTAAGGACCAAGCCTTCCAAAATATTTTAAACCAATATATCGGT	927
Db	841	GCAGCAATATCAAGGTTTAAGGACCAAGCCTTCCAAAATATTTTAAACCAATATATCGGT	900
QY	928	GGTGAGATTCAAAGCTCAAGCAAGTGTACATNTAAACGAA 966	
Db	901	GGTGAGATTCAAAGCTCAAGCAAGTGTACATNTAAACGAA 939	

```

: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al.
: TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
: TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
: FILE REFERENCE: P47H00-07A
: CURRENT APPLICATION NUMBER: US/09/583,110
: CURRENT FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/107,433
: PRIOR FILING DATE: 1998-06-30
: PRIOR APPLICATION NUMBER: US 60/085,131
: PRIOR FILING DATE: 1998-05-12
: PRIOR APPLICATION NUMBER: US 60/051,553
: PRIOR FILING DATE: 1997-07-02
: NUMBER OF SEQ ID NOS: 5342
: SEQ ID NO 2539
: LENGTH: 942
: TYPE: DNA
: ORGANISM: Streptococcus pneumoniae
US-09-583-110-2539

```

Query Match	96.78;	Score 934.2;	DB 25;	Length 942;
Best Local Similarity	99.78;	Pred. No. 2.9e-181;		
Matches 936; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0

OY	28	ATGAAGAAAAAATTTATGTCAGSTGGCATCACACCTTTATACATACCACTTTAGACGCT	87
Db	1	ATGAAGAAAAAATTTATGTCAGSTGGCATCACACCTTTATACATACCACTTTAGACGCT	60
OY	88	TGTTGAAAAGGCTCAGAAAGTGCAGACCTTATCAGCATGAAAGGGAGTGCATTACAGAA	147
Db	61	TGTTGAAAAGGCTCAGAAAGGCGCAGACCTTATCAGCATGAAAGGGAGTGCATTACAGAA	120
OY	148	CATCAATTTTATGAGCAGATGAAAAAGCAACCCCTTCAGCCCAACAACTCTGTAAATATG	207
Db	121	CATCAATTTTATGAGCAGATGAAAAAGCAACCCCTTCAGCCCAACAACTCTGTAAATATG	180
OY	208	ACCATCCAAAAAGTTTTTGAAAAACAAATATGGCTCAGACCTTGATATAAAGATGTGAT	267
Db	181	ACCATCCAAAAAGTTTTTGAAAAACAAATATGGCTCAGACCTTGATATAAAGATGTGAT	240
OY	268	GATATATTGGCGAAGAAAAAAAACAAATATGGCGAAATATACCAAGCTGCTTGTACAA	327
Db	241	GATATATTGGCGAAGAAAAAAAACAAATATGGCGAAATATACCAAGCTGCTTGTACAA	300
OY	328	GCAGGTATGACTCTTTGAAAACACGTAAGCTCAAAATTCGTACAAGTAATTAAGTTGAGTTG	387
Db	301	GCAGGTATGACTCTTTGAAAACACGTAAGCTCAAAATTCGTACAAGTAATTAAGTTGAGTTG	360
OY	388	GCAGTTAAGAAAGGTAGCAAGACGCTAAATGACAGATGAAGCCCTATTAAGAAAGCTTTGAT	447
Db	361	GCAGTTAAGAAAGGTAGCAAGACGCTAAATGACAGATGAAGCCCTATTAAGAAAGCTTTGAT	420
OY	448	GAGTACACTCCAGATGTAAAGGCTCAAAATCATCCGCTTAAATATGAAATAGAAAGGCCAA	507
Db	421	GAGTACACTCCAGATGTAAAGGCTCAAAATCATCCGCTTAAATATGAAATAGAAAGGCCAA	480
OY	508	GAAGTTCGCAAAAAGCCAGGCAAGAGTGTGATTTGGCTCAATTAAGCCAAAGATAT	567
Db	481	GAAGTTCGCAAAAAGCCAGGCAAGAGTGTGATTTGGCTCAATTAAGCCAAAGATAT	540
OY	568	TCAACTGTGAAAAAACAAGAAATGTTGATGAGAAATTAACCTTGTATCTGCTTCACAA	627
Db	541	TCAACTGTGAAAAAACAAGAAATGTTGATGAGAAATTAACCTTGTATCTGCTTCACAA	600
OY	628	GAAGTACTGAGCAAGTCAAAAAAGCCGCTTTCGCTTATAGATGTGATGGTGTTCGTAT	687
Db	601	GAAGTACTGAGCAAGTCAAAAAAGCCGCTTTCGCTTATAGATGTGATGGTGTTCGTAT	660
OY	688	GTGATTAACAGCAACTGCGACACAAAGCTTACAGTATACATATTAATTTAAACATCAAT	747
Db	661	GTGATTAACAGCAACTGCGACACAAAGCTTACAGTATACATATTAATTTAAACATCAAT	720
OY	748	AAGAAAACAGAAAAATCATCTTAATTTGATGACTACAAAGAAAAATTAAGAAAGCTGTATC	807

Db	721	AAGAAAACGAAAAATTCATCTATATATGATGATCAAAAAGAAAAATTAATAAACTGTATC	780
Qy	808	TTGACTCAAAAACCAAAATGATTCACATTTGTTCAAAGCATTAATCGSAAAAGAAATGCAA	867
Db	761	TTGACTCAAAAACCAAAATGATTCACATTTGTTCAAAGCATTAATCGSAAAAGAAATGCAA	840
Qy	868	GCAGCCATATATCAAGTTAAGSACCAAGCCTTCAAAATATCTTTACCCAAATATACGGT	927
Db	841	GCAGCCATATATCAAGTTAAGSACCAAGCCTTCAAAATATCTTTACCCAAATATACGGT	900
Qy	928	GSTGAGATTCAGCTCAAGCAGCTGTGATNATCAAAACGAA	966
Db	901	GSTGAGATTCAGCTCAAGCAGCTGTGATNATCAAAACGAA	939

RESULT 7
US-09-107-433-1524

; Sequence 1524, Application US/091074333
; GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
SEQUENCES RELATING TO CRYPTOSPORIDIUM

SEQUENCES REMAINING IN SINGLETONS PNEUMONIA FOR DIA
THERAPEUTICS
NUMBER OF SEQUENCES: 5206

;; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street
CITY: Waltham

STATE: MASS
COUNTRY: USA
ZIP: 02354

DATE: 0209
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660

```
;
;
COMPUTER: <Unknown>
OPERATING SYSTEM: <Unknown>
```

```

; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/09/107,433
FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/ 085131
FILING DATE: May 12, 1998

APPLICATION NUMBER: 60/051553
FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:
NAME: Arinello, Pamela Denek

REGISTRATION NUMBER: 40,
REFERENCE/DOCKET NUMBER:

TELECOMMUNICATION INFORMATION
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8327

TELEFAX: (701) 855-8211
; INFORMATION FOR SEQ ID NO: 1524:
; SEQUENCE CHARACTERISTICS:

LENGTH: 888 base pairs
TYPE: nucleic acid

```

; STRANDEDNESS: double
; TOPOLOGY: circular

```

MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO

ANAL-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: *Streptococcus pneumoniae*

```

/ ORGANISM: Streptococcus pneumoniae
;
; FEATURE:
; NAME/KEY: misc feature

```

```

; LOCATION: (B) LOCATION 1...888
; SEQUENCE DESCRIPTION: SEQ ID NO: 1524:

```

US-09-107-433-1524

Query Match	89.98%	Score 868.2	DB 16	Length 888
Best Local Similarity	99.58%	Pred. No. 9,9e-168		
Matches 881	Conservative	0	Mismatches 3	Indels 1
				Gaps 1

```

QY 83 CAGCTTGTTCAGAAAGG-GTCAGAAAGGTGCAGACCTTATCAGCATGAAAGGGAGTGCATT 141
DB 1 CAGCTTGTTCAGAAAGG-GTCAGAAAGGTGCAGACCTTATCAGCATGAAAGGGAGTGCATT 60
QY 142 ACAGAACATCAATTTTATGAGCAAGTGAAGCAACCTTCAGCCCAACAGTCTGTTA 201
DB 61 ACAGAACATCAATTTTATGAGCAAGTGAAGCAACCTTCAGCCCAACAGTCTGTTA 120
QY 202 AATATGACCATCCAAAAGTTTGTGAAAACCAATATGCTCAGAGCTGTATGATTAAGAG 261
DB 121 AATATGACCATCCAAAAGTTTGTGAAAACCAATATGCTCAGAGCTGTATGATTAAGAG 180
QY 262 GTTGATGATATCTATGCGGCAAGAAAACCAATATGCGGCAAGAAAACCAAGCTGTCTTG 321
DB 181 GTTGATGATATCTATGCGGCAAGAAAACCAATATGCGGCAAGAAAACCAAGCTGTCTTG 240
QY 322 TCACAAAGCAGTATGACTCTTGAACACGTAAGCTCAATTTGGTACAAATTAATTTGTT 381
DB 241 TCACAAAGCAGTATGACTCTTGAACACGTAAGCTCAATTTGGTACAAATTAATTTGTT 300
QY 382 GAGTTGCACTTAAGAAAGTAGAGAGAGTGAATTCAGATGAAGCTTAAAGAAAGCC 441
DB 301 GAGTTGCACTTAAGAAAGTAGAGAGAGTGAATTCAGATGAAGCTTAAAGAAAGCC 360
QY 442 TTGATGATGATACCTCCAGATGTACGCGCTCAATCCGCTTAAATTAAGATTAAG 501
DB 361 TTGATGATGATACCTCCAGATGTACGCGCTCAATCCGCTTAAATTAAGATTAAG 420
QY 502 GCCAAAGAAAGTCTCCGAAAAGCCAGAGGAGGCTGATTTTGTCAATTAAGCCAA 561
DB 421 GCCAAAGAAAGTCTCCGAAAAGCCAGAGGAGGCTGATTTTGTCAATTAAGCCAA 480
QY 562 GATATTCACACTGATGAAAACCAAAAGAAATGTTGGAGAAATTAACCTTGTATTCGT 621
DB 481 GATATTCACACTGATGAAAACCAAAAGAAATGTTGGAGAAATTAACCTTGTATTCGT 540
QY 622 TCACAAAGCAGTATGACTCTTGAACACGTAAGCTCAATTTGGTACAAATTAATTTGTT 681
DB 541 TCACAAAGCAGTATGACTCTTGAACACGTAAGCTCAATTTGGTACAAATTAATTTGTT 600
QY 682 TCGATGATGATACCTCCAGATGTACGCGCTCAATCCGCTTAAATTAAGATTAAG 741
DB 601 TCGATGATGATACCTCCAGATGTACGCGCTCAATCCGCTTAAATTAAGATTAAG 660
QY 742 CTCACATGAGAAAACAGAAAATCATCTAATATTTGATGATGATGATGATGATGATGAT 801
DB 661 CTCACATGAGAAAACAGAAAATCATCTAATATTTGATGATGATGATGATGATGATGAT 720
QY 802 GTTATCTTGAAGCAAGCAAAATGATTCACATTTGTTCAAGCATTAATCGGAAAAGAA 861
DB 721 GTTATCTTGAAGCAAGCAAAATGATTCACATTTGTTCAAGCATTAATCGGAAAAGAA 780
QY 862 TTGCAAGCAGCCTATATCAAGGTTAAGAGCAGCCTTCCAAAATATCTTTACCAATAT 921
DB 781 TTGCAAGCAGCCTATATCAAGGTTAAGAGCAGCCTTCCAAAATATCTTTACCAATAT 840
QY 922 ATCGGTGATGAGATTCAGCTCAAGCAGTATGATGATGATGATGATGATGATGATGAT 966
DB 841 ATCGGTGATGAGATTCAGCTCAAGCAGTATGATGATGATGATGATGATGATGATGATGAT 885

```

RESULT 8

US-60-061-998-559

; Sequence 559, Application US/60061998
; GENERAL INFORMATION:

; APPLICANT: LAGACE, ROBERT E.
; APPLICANT: CORLEY, NEIL C.
; APPLICANT: RUSSO, FRANK D.
; APPLICANT: HANN, AMY L.
; APPLICANT: HEATH, JOE D.
; APPLICANT: FINNEY, GREGORY L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF STREPTOCOCCUS PNEUMONIAE
; TITLE OF INVENTION: GENOME, FRAGMENTS THEREOF, AND USES THEREOF

```

? NUMBER OF SEQUENCES: 797
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
? STREET: 3174 PORTER DRIVE
? CITY: PALO ALTO
? STATE: CALIFORNIA
? COUNTRY: USA
? ZIP: 94304
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/60/061,998
? FILING DATE:
? CLASSIFICATION:
? ATTORNEY/AGENT INFORMATION:
? NAME: CERRONE, MICHAEL C.
? REGISTRATION NUMBER: 39,132
? REFERENCE/DOCKET NUMBER: PM-0006-2P
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (650) 855-0555
? TELEFAX: (650) 845-4166
? INFORMATION FOR SEO ID NO: 559:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 11591 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? IMMEDIATE SOURCE:
? CLONE: SPN1564
? US-60-061-998-559

Query Match 83.68; Score 807.8; DB 60; Length 11591;
Best Local Similarity 98.98; Pred. No. 4.5e-155;
Matches 824; Conservative 0; Mismatches 7; Indels 2; Gaps 1;

QY 1 AGTACACTTATCTCAAGAGAGTAGACATGAGAAAATTAATTTGGCAGGTGCATCACA 60
DB 6276 AGTACACTTATCTCAAGAGAGTAGACATGAGAAAATTAATTTGGCAGGTGCATCACA 6335
QY 61 CTATTATCAGTAGACACTTTAGCAGCTTGTTCGAAAAGGTGAGAGGTGACACTTATC 120
DB 6336 CTATTATCAGTAGACACTTTAGCAGCTTGTTCGAAAAGGTGAGAGGTGACACTTATC 6395
QY 121 AGCATGAAGGGGATGTCATTACAGAACATCAATTTTATGAGCAAGTGAAGCAACCT 180
DB 6396 AGCATGAAGGGGATGTCATTACAGAACATCAATTTTATGAGCAAGTGAAGCAACCT 6455
QY 181 TCAGCCCAAGAGCTGTTAAATATGACATCCAAAAGTTTGTGAAAACCAATATGAGC 240
DB 6456 TCAGCCCAAGAGCTGTTAAATATGACATCCAAAAGTTTGTGAAAACCAATATGAGC 6515
QY 241 TCAGAGCTGATGATTAAGAGGTGATGATGATGATGATGATGATGATGATGATGATGAT 300
DB 6516 TCAGAGCTGATGATTAAGAGGTGATGATGATGATGATGATGATGATGATGATGATGAT 6575
QY 301 GAAACTACCAAGCTGCTTGTCAAGCAGATGATGACTCTTGAACAGCTAAAGCTCAA 360
DB 6576 GAAACTACCAAGCTGCTTGTCAAGCAGATGATGACTCTTGAACAGCTAAAGCTCAA 6635
QY 361 ATTGCTACAGTAAATAGTTGAGTGGCAGTTAAGAGGTAGAGAGGTGAATGACA 420
DB 6636 ATTGCTACAGTAAATAGTTGAGTGGCAGTTAAGAGGTAGAGAGGTGAATGACA 6695
QY 421 GATGAAGCTTATGAAGAGCTTTGATGATGATGATGATGATGATGATGATGATGATGAT 478
DB 6696 GATGAAGCTTATGAAGAGCTTTGATGATGATGATGATGATGATGATGATGATGATGAT 6755
QY 479 TCGCTCTTAATTAAGAAATGAAGCCAAAGAGTTCTGAAAAAGCCAAAGCAGAGGTG 538
DB TCGCTCTTAATTAAGAAATGAAGCCAAAGAGTTCTGAAAAAGCCAAAGCAGAGGTG 538

```


TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 33:

SEQUENCE CHARACTERISTICS:
 LENGTH: 640 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 33:
 US-09-765-271-33

Query Match 57.2%; Score 552.4; DB 33; Length 640;
 Best Local Similarity 94.7%; Pred. No. 5.4e-103;
 Matches 595; Conservative 0; Mismatches 26; Indels 7; Gaps 2;

QY 90 TTGGAAGGCTCAGAGGTGACAGCTTATCAGCATGAAAGGGGATGCTATTACAGACA 149
 DB 1 TTGGAAGGCTCAGAGGTGACAGCTTATCAGCATGAAAGGGGATGCTATTACAGACA 60
 QY 150 TCAATTTTATGACAGTGAAGCAACCTTCAGCCCAACAGCTCTGTTAATATGAC 209
 DB 61 TCATTTTATGACAGTGAAGCAACCTTCAGCCCAACAGCTCTGTTAATATGAC 120
 QY 210 CATCAAAAAGTTTGAAGCAATATGCTCAGAGCTTGATGATTAAGAGTTGATGA 269
 DB 121 CATCAAAAAGTTTGAAGCAATATGCTCAGAGCTTGATGATTAAGAGTTGATGA 180
 QY 270 TACTATTGCCGAAGAAAAACAATATGCGGAAAACTACCAAGCTGCTTGACAAAGC 329
 DB 181 TACTATTGCCGAAGAAAAACAATATGCGGAAAACTACCAAGCTGCTTGACAAAGC 240
 QY 330 AGGTATGACTCTGGAAGACGTAAAGCTCAATTCGACAAAGTAAATAGTTAGTTGGC 389
 DB 241 AGGTATGACTCTGGAAGACGTAAAGCTCAATTCGACAAAGTAAATAGTTAGTTGGC 300
 QY 390 AGTTAAGAGTGAAGCAAGCTGAATGACATGAAGCTATAGAAAGCTTTGATGA 449
 DB 301 AGTTAAGAGTGAAGCAAGCTGAATGACATGAAGCTATAGAAAGCTTTGATGA 360
 QY 450 GTACACTCCAGATGTAAGCGCTCAATTCCTTTAATATGAGATTAAGGCCAAAGA 509
 DB 361 GTACACTCCAGATGTAAGCGCTCAATTCCTTTAATATGAGATTAAGGCCAAAGA 420
 QY 510 AGTTCTGAAAAAGCAAGGACAGAGTGTGATTTTGTCAATTAAGCAAGATAATTC 569
 DB 421 AGTTCTGAAAAAGCAAGGACAGAGTGTGATTTTGTCAATTAAGCAAGATAATTC 480
 QY 570 AACTGATGAAAAAACAAGAAATGTTGAGAAATTAACCTTGATTCTGCTTCAACAGA 629
 DB 481 AACTGATGAAAAAACAAGAAATGTTGAGAAATTAACCTTGATTCTGCTTCAACAGA 540
 QY 630 AGTACTG--AGCAAGTCAAAAAAGCGCTTTCGCTTTAGATGATGATGATGAT 687
 DB 541 AGTACTGAGCAAGTCAAAAAAGCGCTTTCGCTTTAGATGATGATGATGATGAT 600
 QY 688 G-----TGATTACAGCACTGACACAGA 710
 DB 601 GATGTGATTAACAGCACTGACACAGA 628

RESULT 11

US-09-765-272-33
 Sequence 33, Application US/09765272
 GENERAL INFORMATION:

APPLICANT: Choi et. al.
 TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
 NUMBER OF SEQUENCES: 452
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland

COUNTRY: USA
 ZIP: 20850
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/765, 272
 FILING DATE: 22-Jan-2001
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/961,083
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Brookes, A. Anders
 REGISTRATION NUMBER: 36,373
 REFERENCE/DOCKET NUMBER: PB340P2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 33:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 640 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 33:
 US-09-765-272-33

Query Match 57.2%; Score 552.4; DB 33; Length 640;
 Best Local Similarity 94.7%; Pred. No. 5.4e-103;
 Matches 595; Conservative 0; Mismatches 26; Indels 7; Gaps 2;

QY 90 TTGGAAGGCTCAGAGGTGACAGCTTATCAGCATGAAAGGGGATGCTATTACAGACA 149
 DB 1 TTGGAAGGCTCAGAGGTGACAGCTTATCAGCATGAAAGGGGATGCTATTACAGACA 60
 QY 150 TCAATTTTATGACAGTGAAGCAACCTTCAGCCCAACAGCTCTGTTAATATGAC 209
 DB 61 TCATTTTATGACAGTGAAGCAACCTTCAGCCCAACAGCTCTGTTAATATGAC 120
 QY 210 CATCAAAAAGTTTGAAGCAATATGCTCAGAGCTTGATGATTAAGAGTTGATGA 269
 DB 121 CATCAAAAAGTTTGAAGCAATATGCTCAGAGCTTGATGATTAAGAGTTGATGA 180
 QY 270 TACTATTGCCGAAGAAAAACAATATGCGGAAAACTACCAAGCTGCTTGACAAAGC 329
 DB 181 TACTATTGCCGAAGAAAAACAATATGCGGAAAACTACCAAGCTGCTTGACAAAGC 240
 QY 330 AGGTATGACTCTGGAAGACGTAAAGCTCAATTCGACAAAGTAAATAGTTAGTTGGC 389
 DB 241 AGGTATGACTCTGGAAGACGTAAAGCTCAATTCGACAAAGTAAATAGTTAGTTGGC 300
 QY 390 AGTTAAGAGTGAAGCAAGCTGAATGACATGAAGCTATAGAAAGCTTTGATGA 449
 DB 301 AGTTAAGAGTGAAGCAAGCTGAATGACATGAAGCTATAGAAAGCTTTGATGA 360
 QY 450 GTACACTCCAGATGTAAGCGCTCAATTCCTTTAATATGAGATTAAGGCCAAAGA 509
 DB 361 GTACACTCCAGATGTAAGCGCTCAATTCCTTTAATATGAGATTAAGGCCAAAGA 420
 QY 510 AGTTCTGAAAAAGCAAGGACAGAGTGTGATTTTGTCAATTAAGCAAGATAATTC 569
 DB 421 AGTTCTGAAAAAGCAAGGACAGAGTGTGATTTTGTCAATTAAGCAAGATAATTC 480
 QY 570 AACTGATGAAAAAACAAGAAATGTTGAGAAATTAACCTTGATTCTGCTTCAACAGA 629
 DB 481 AACTGATGAAAAAACAAGAAATGTTGAGAAATTAACCTTGATTCTGCTTCAACAGA 540
 QY 630 AGTACTG--AGCAAGTCAAAAAAGCGCTTTCGCTTTAGATGATGATGATGAT 687
 DB 541 AGTACTGAGCAAGTCAAAAAAGCGCTTTCGCTTTAGATGATGATGATGATGAT 600

Query Match 27.5%; Score 265.6; DB 13; Length 451;
Best Local Similarity 97.9%; Pred. No. 3.2e-44;
Matches 279; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 1 AGTAACTTATCTCAAGAGTAGACATGAGAAAAAATTTATGGCAGGTGCATCACA 60
DB 167 AGTAACTTATCTCAAGAGTAGACATGAGAAAAAATTTATGGCAGGTGCATCACA 226
QY 61 CTAATTACATAGCACTTTAGCAGCTTGTTCGAAAGGTCAGAGGTGCAGACTTATC 120
DB 227 CTAATTACATAGCACTTTAGCAGCTTGTTCGAAAGGTCAGAGGTGCAGACTTATC 286
QY 121 AGCATGAAGGGATGTCTTACAGAACATCAATTTATGAGCAAGTGAAGCAACCT 180
DB 287 AGCATGAAGGGATGTCTTACAGAACATCAATTTATGAGCAAGTGAAGCAACCT 346
QY 181 TCAGCCCAACAGCTTGTAAATATGACCATCCAAAAGTTTGG-AAAAACAATATGG 239
DB 347 TCAGCCCAACAGCTTGTAAATATGACCATCCAAAAGTTTGG-AAAAACAATATGG 406
QY 240 CTCAGAGCTTGATGATTAAGAGGTGATGATGATGATGATGATGATGATGATG 284
DB 407 CTCAGAGCTTGATGATTAAGAGGTGATGATGATGATGATGATGATGATGATG 451

RESULT 14

PCT-US97-07950-219
Sequence 219, Application PC/TUS9707950

GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Nicholas, Richard
APPLICANT: Stodola, Robert
TITLE OF INVENTION: Novel Compounds
NUMBER OF SEQUENCES: 552
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/07950
FILING DATE: 14-MAY-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/017670
FILING DATE: 14-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmil, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50475
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 219:
SEQUENCE CHARACTERISTICS:
LENGTH: 791 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
PCT-US97-07950-219

Query Match 27.5%; Score 265.6; DB 1; Length 791;
Best Local Similarity 97.9%; Pred. No. 3.7e-44;
Matches 279; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 1 AGTAACTTATCTCAAGAGTAGACATGAGAAAAAATTTATGGCAGGTGCATCACA 60
DB 507 AGTAACTTATCTCAAGAGTAGACATGAGAAAAAATTTATGGCAGGTGCATCACA 566
QY 61 CTAATTACATAGCACTTTAGCAGCTTGTTCGAAAGGTCAGAGGTGCAGACTTATC 120
DB 567 CTAATTACATAGCACTTTAGCAGCTTGTTCGAAAGGTCAGAGGTGCAGACTTATC 626
QY 121 AGCATGAAGGGATGTCTTACAGAACATCAATTTATGAGCAAGTGAAGCAACCT 180
DB 627 AGCATGAAGGGATGTCTTACAGAACATCAATTTATGAGCAAGTGAAGCAACCT 686
QY 181 TCAGCCCAACAGCTTGTAAATATGACCATCCAAAAGTTTGG-AAAAACAATATGG 239
DB 687 TCAGCCCAACAGCTTGTAAATATGACCATCCAAAAGTTTGG-AAAAACAATATGG 746
QY 240 CTCAGAGCTTGATGATTAAGAGGTGATGATGATGATGATGATGATGATGATG 284
DB 747 CTCAGAGCTTGATGATTAAGAGGTGATGATGATGATGATGATGATGATGATG 791

RESULT 15

PCT-US97-07950-219
Sequence 219, Application PC/TUS9707950

GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Nicholas, Richard
APPLICANT: Stodola, Robert
TITLE OF INVENTION: Novel Compounds
NUMBER OF SEQUENCES: 552
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/07950
FILING DATE: 14-MAY-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/017670
FILING DATE: 14-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmil, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50475
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 219:
SEQUENCE CHARACTERISTICS:
LENGTH: 791 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
PCT-US97-07950-219

Query Match 27.5%; Score 265.6; DB 2; Length 791;

Best Local Similarity 97.98; Pred. No. 3.7e-44;
Matches 279; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY	1	AGTAACACTTATCTCAAGAGAGTAGACATGAAGAAATTTATGTCAGAGTGCCATCACA	60
Db	507	AGTAACACTTATCTCAAGAGAGTAGACATGAAGAAATTTATGTCAGAGTGCCATCACA	566
QY	61	CTATTATCAGTAGCAACTTTAGCAGCTTGTGAAAAGGTCAGAAAGTGCAGACCTTATC	120
Db	567	CTATTATCAGTAGCAACTTTAGCAGCTTGTGAAAAGGTCAGAAAGGTCAGACCTTATC	626
QY	121	AGCATGAAGGGGATGTCTATACAGACATCAATTTATGAGCAAGTGAAAACCAACCT	180
Db	627	AGCATGAAGGGGATGTCTATACAGACATCAATTTATGAGCAAGTGAAAACCAACCT	686
QY	181	TCAGCCCAACAAGTCTTGTAAATATGACATCCAAAAAGTTTG-AAAAACATATG	239
Db	687	TCAGCCCAACAAGTCTTGTAAATATGACATCCAAAAAGTTTGAAAAAACATATG	746
QY	240	CTCAGAGCTTGATGATGAAGAGTGTGATGATCTATTCGCGAGA	284
Db	747	CTCAGAGCTTGATGATGAAGAGTGTGATGATCTATTCGCGAGA	791

Search completed: September 4, 2003, 21:52:47
Job time : 2930 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2003, 20:27:27 ; Search time 187 Seconds
(without alignments)
5693.830 Million cell updates/sec

Title: US-10-049-473a-1
Perfect score: 966
Sequence: 1 agtaacctatctcaag.....gcagtagcatcaacgaag 966

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 907674 seqs, 55110846 residues

Total number of hits satisfying chosen parameters: 1815348

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries:

Database : Pending_Patents_NA_New:*
1: /cgn2_6/pdata/2/pna/PC7_NEW_COMB.seq:*
2: /cgn2_6/pdata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/pdata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/pdata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/pdata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/pdata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/pdata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	934.2	96.7	942	US-10-640-833-2539	Sequence 2539, Ap
2	868.2	89.9	888	US-10-617-320-1524	Sequence 1524, Ap
3	52	5.4	1107	US-10-603-113-5692	Sequence 5692, Ap
4	51.4	5.3	672	US-10-603-113-7512	Sequence 7512, Ap
5	46.2	4.8	1263	US-10-602-494-313	Sequence 1842, Ap
6	45.2	4.7	3095	US-10-602-494-313	Sequence 313, App
7	44.6	4.6	17703	US-10-257-1668-34	Sequence 34, App
8	44.4	4.6	1287	US-10-603-113-5698	Sequence 3698, Ap
9	44.2	4.6	342	US-10-603-113-5931	Sequence 5931, Ap
10	43.8	4.5	96596	PCR-US02-38582-100	Sequence 100, App
11	43.4	4.5	3231	US-10-257-1668-164	Sequence 164, App
12	43.2	4.5	653	US-10-626-717-3404	Sequence 3404, Ap
13	43	4.5	2250	US-10-602-494-302	Sequence 302, App
14	43	4.5	5898	US-10-640-833-1168	Sequence 1168, Ap
15	43	4.5	5919	US-10-617-320-648	Sequence 648, App
16	43	4.5	50000	US-09-662-2548-23	Sequence 23, App
17	42.2	4.4	891	US-10-603-113-11247	Sequence 11247, A
18	42.2	4.4	1835	US-10-612-783-2093	Sequence 2093, Ap
19	42.2	4.4	2754	US-10-603-113-4857	Sequence 4857, Ap
20	42.2	4.4	199720	US-10-257-1668-5	Sequence 5, Appl
21	42	4.3	17220	US-10-602-494-327	Sequence 327, App
22	41.8	4.3	7892	US-10-257-1668-138	Sequence 138, App
23	41.6	4.3	2082	US-10-603-113-4064	Sequence 4064, Ap
24	41.2	4.3	40178	US-10-600-009-467	Sequence 467, App
25	41.2	4.3	40178	US-10-600-009-468	Sequence 468, App
26	41.2	4.3	50423	US-60-485-450-12101	Sequence 12101, A

27	41	4.2	453	6	US-10-640-833-1702	Sequence 1702, Ap
28	41	4.2	477	6	US-10-617-320-1055	Sequence 1055, Ap
29	41	4.2	1367	6	US-10-302-172-553	Sequence 553, App
30	40.8	4.2	1785	6	US-10-603-113-928	Sequence 928, App
31	40.8	4.2	1791310	5	US-09-947-914-46	Sequence 46, Appl
32	40.6	4.2	5271	6	US-10-450-826-125	Sequence 125, App
33	40.6	4.2	22268	6	US-10-450-826-126	Sequence 126, Appl
34	40.4	4.2	3463	6	US-10-299-636-93	Sequence 93, Appl
35	40.4	4.2	17220	6	US-10-602-494-185	Sequence 185, App
36	40.2	4.2	3561	6	US-10-603-113-2080	Sequence 2080, Ap
37	40.2	4.2	11473	6	US-10-257-1668-111	Sequence 111, App
38	40	4.1	2377	6	US-10-602-494-231	Sequence 231, App
39	40	4.1	2469	6	US-10-603-113-4746	Sequence 4746, Ap
40	39.8	4.1	2229	6	US-10-602-494-240	Sequence 240, App
41	39.8	4.1	2382	6	US-10-602-494-325	Sequence 325, App
42	39.6	4.1	1269	6	US-10-603-113-1601	Sequence 1601, Ap
43	39.6	4.1	7441	6	US-10-257-1668-140	Sequence 140, App
44	39.6	4.1	7615	6	US-10-257-1668-53	Sequence 53, Appl
45	39.6	4.1	171596	7	US-60-487-610-19488	Sequence 19488, A

ALIGNMENTS

RESULT 1
US-10-640-833-2539
Sequence 2539, Application US/10640833
GENERAL INFORMATION:
APPLICANT: Doucette-Stamm, Lynn
APPLICANT: Bush, David
APPLICANT: Zeng, Qiantong
APPLICANT: Opperman, Timothy
APPLICANT: Houseweart, Chad Eric
TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
FILE REFERENCE: PAT03-15
CURRENT APPLICATION NUMBER: US/10/640, 833
CURRENT FILING DATE: 2003-08-14
PRIOR APPLICATION NUMBER: US 09/583, 110
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/107, 433
PRIOR FILING DATE: 1998-06-30
PRIOR APPLICATION NUMBER: US 60/085, 131
PRIOR FILING DATE: 1998-05-12
PRIOR APPLICATION NUMBER: US 60/051, 553
PRIOR FILING DATE: 1997-07-02
NUMBER OF SEQ ID NOS: 5322
SEQ ID NO 2539
LENGTH: 942
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
US-10-640-833-2539

Query Match 96.7%; Score 934.2; DB 6; Length 942;
Best Local Similarity 99.7%; Pred. No. 3.7e-247;
Matches 936; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY	28	ATGAAGAAAAATATATGTCAGTGCATCAGCACTTTATGCTAGCACTTAGAGCT	87
DB	1	ATGAAGAAAAATATATGTCAGTGCATCAGCACTTTATGCTAGCACTTAGAGCT	60
QY	88	TGTTCAAGAGGTCAGCAAGTGCAGCACTTATCAGCATGAAGGGATCTCATACAGAA	147
DB	61	TGTTCAAGAGGTCAGCAAGGTCAGCACTTATCAGCATGAAGGGATCTCATACAGAA	120
QY	148	CATCAATTTATGAGCAAGTGAAGCAACCTTCAGCCCAACAACTCTGTTAAATATG	207
DB	121	CATCAATTTATGAGCAAGTGAAGCAACCTTCAGCCCAACAACTCTGTTAAATATG	180
QY	208	ACCATCAAAAGTTTGAAGCAATATGCTAGCACTTGATGATTAAGGTTGAT	267
DB	181	ACCATCAAAAGTTTGAAGCAATATGCTAGCACTTGATGATTAAGGTTGAT	240

QY	268	GATACATATTCGCGAAGAAAAAACAATATGGCGAAACTACCAACAGCTGCTTGTACAA	327
Db	241	GATACATATTCGCGAAGAAAAAACAATATGGCGAAACTACCAACAGCTGCTTGTACAA	300
QY	328	GCAGGTATGACTCTTGAAACACGTAAAGCTCAAAATCTACAAAGTAAATAGTGGTTG	387
Db	301	GCAGGTATGACTCTTGAAACACGTAAAGCTCAAAATCTGACAAAGTAAATAGTGGTTG	360
QY	388	GCAGTTAAGAAGGTGCGCAAGCTGAATTGACAGATGAAGCCTATPAGAAAGCTTTGAT	447
Db	361	GCAGTTAAGAAGGTGCGCAAGCTGAATTGACAGATGAAGCCTATPAGAAAGCCTTTGAT	420
QY	448	GAGTACACTCCAGATGTAAACGGCTCAAATCATCCGCTCTTAATATGAAGATTAAGGCCAA	507
Db	421	GAGTACACTCCAGATGTAAACGGCTCAAATCATCCGCTCTTAATATGAAGATTAAGGCCAA	480
QY	508	GAACTTCTCGAAAAAGCCAGGCAAGAGGTGCTGATTTTGCTCAATTAGCCAAAGATAT	567
Db	481	GAACTTCTCGAAAAAGCCAGGCAAGAGGTGCTGATTTTGCTCAATTAGCCAAAGATAT	540
QY	568	TCACTGATGAAAAAACAATAAGTGGTGAGAAATTACCTTGATTCGCTCCACCA	627
Db	541	TCACTGATGAAAAAACAATAAGTGGTGAGAAATTACCTTGATTCGCTCCACCA	600
QY	628	GAACTACTGAGCAAGTCAAAAAAGCCGCTTCGCTTAGATGTGATGTGTTCTGAT	687
Db	601	GAACTACTGAGCAAGTCAAAAAAGCCGCTTCGCTTAGATGTGATGTGTTCTGAT	660
QY	688	GTGATTAAGCAACGTGGCACAACGCTACAGTACCAATATTTACATTTGTAAACTACT	747
Db	661	GTGATTAAGCAACGTGGCACAACGCTACAGTACCAATATTTACATTTGTAAACTACT	720
QY	748	AAGAAAACAGAAAAATCATCTAATTTGATGACTCAAGAAGAAAAATPAAAACTGTATC	807
Db	721	AAGAAAACAGAAAAATCATCTAATTTGATGACTCAAGAAGAAAAATPAAAACTGTATC	780
QY	808	TTGACTCAAAAACAAAATGATTCACAATTTGTTCAAAGCATTTATCGAAAAAGATTGCAA	867
Db	781	TTGACTCAAAAACAAAATGATTCACAATTTGTTCAAAGCATTTATCGAAAAAGATTGCAA	840
QY	868	GCAGCCCATATCAAGGTTAAGGACCAAGCCTCCCAAAATATCTTTACCAATATATCGGT	927
Db	841	GCAGCCCATATCAAGGTTAAGGACCAAGCCTCCCAAAATATCTTTACCAATATATCGGT	900
QY	928	GGTGAGATTCAAACTCAAGCAGTAGTACATCAAAACGAA	966
Db	901	GGTGAGATTCAAACTCAAGCAGTAGTACATCAAAACGAA	939

RESULT 2
 US-10-617-320-1524
 Sequence 1524, Application US/10617320
 GENERAL INFORMATION:
 APPLICANT: Lynn A Doucette-Stamm and David Bush
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
 SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
 THERAPEUTICS
 NUMBER OF SEQUENCES: 5206
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: GENOME THERAPEUTICS CORPORATION
 STREET: 100 Beaver Street
 City: Waltham
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02354
 COMPUTER READABLE FORM:
 MEDIUM TYPE: CD/ROM ISO9660
 COMPUTER: <Unknown>
 OPERATING SYSTEM: <Unknown>
 SOFTWARE: <Unknown>
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/617,320
 FILING DATE: 10-Jul-2003

```

1      PRIOR APPLICATION DATA:
2
3      APPLICATION NUMBER: US/09/107,433
4
5      FILING DATE: 30-Jun-1998
6
7      APPLICATION NUMBER: 60/ 085131
8
9      FILING DATE: May 12, 1998
10
11     APPLICATION NUMBER: 60/051553
12
13     FILING DATE: July 2, 1997
14
15     ATTORNEY/AGENT INFORMATION:
16
17     NAME: Arinello, Pamela Deneke
18
19     REGISTRATION NUMBER: 40,489
20
21     REFERENCE/DOCKET NUMBER: GTC-011
22
23     TELECOMMUNICATION INFORMATION:
24
25     TELEPHONE: (781)893-5007
26
27     TELEFAX: (781)893-8277
28
29     INFORMATION FOR SEQ ID NO: 1524:
30
31     SEQUENCE CHARACTERISTICS:
32
33     LENGTH: 888 base pairs
34
35     TYPE: nucleic acid
36
37     STRANDEDNESS: double
38
39     TOPOLOGY: circular
40
41     MOLECULE TYPE: DNA (genomic)
42
43     HYPOTHETICAL: NO
44
45     ANTI-SENSE: NO
46
47     ORIGINAL SOURCE:
48
49     ORGANISM: Streptococcus pneumoniae
50
51     FEATURE:
52
53     NAME/KEY: misc.feature
54
55     LOCATION: (B) LOCATION 1...888
56
57     SEQUENCE DESCRIPTION: SEQ ID NO: 1524:
58
59     /S-10-617-320-1524

```

Query Match	89.9%	Score 868.2	DB 6	Length 888
Best Local Similarity	99.5%	Pred. No. 5.1e-229		
Matches 881	Conservative 0	Mismatches 3	Indels 1	Gaps 1
QY	83	CAGCTTGTTCGAAGG-GTCGAAAGGTGCAGACCTTATTCAGCATGGAAGGGAGTGCATT	141	
Db	1	CAGCTTGTTCGAAGGTGTCCGAAAGGGGGCAGACCTTATTCAGCATGGAAGGGAGTGCATC	60	
QY	142	ACGAAACATCATTTTATATGACGACGATGAAGCAACCTTCACCCCAACAGCTTGTTA	201	
Db	61	ACGAAACATCATTTTATATGACGATGAAGCAACCTTCAGCCCAACAGCTTGTTA	120	
QY	202	AATATGACCATTCGAAAAAGTTTTGAAAAACAATATGCTCAGAGCTTGATGATAAG	261	
Db	121	AATATGACCATTCGAAAAAGTTTTGAAAAACAATATGCTCAGAGCTTGATGATAAG	180	
QY	262	GTTGATGATACTATTGGCGAAGAAAAAAACAATATGGCGAAAACTACCAAGCTGCTG	321	
Db	181	GTTGATGATACTATTGGCGAAGAAAAAAACAATATGGCGAAAACTACCAAGCTGCTG	240	
QY	322	TCACAGCAGGTATGACTCTTGAAACACGTAAAGCTCAATTTGTCACAGTAAATTAAGT	381	
Db	241	TCACAGCAGGTATGACTCTTGAAACACGTAAAGCTCAATTTGTCACAGTAAATTAAGT	300	
QY	382	GAGTTGGCAGTTAAGAAAGTGCAGGAAGCTGAATTACAGATGAAGCCATTAAGAAACC	441	
Db	301	GAGTTGGCAGTTAAGAAAGTGCAGGAAGCTGAATTACAGATGAAGCCATTAAGAAACC	360	
QY	442	TTTGATGAGTACACTCCAGATGTAAAGGCTCAAAATCATCCGCTTAATATATGAAGATTAAG	501	
Db	361	TTTGATGAGTACACTCCAGATGTAAAGGCTCAAAATCATCCGCTTAATATATGAAGATTAAG	420	
QY	502	GCCAAAGAGTTCTCGAAAAAGCCCAAGCGCAGAAAGTGTGATTTTGTCTCATTAAGCCAA	561	
Db	421	GCCAAAGAGTTCTCGAAAAAGCCCAAGCGCAGAAAGTGTGATTTTGTCTCATTAAGCCAA	480	
QY	562	GATAATTCACACTGATGAAAAAACAAAAAGAAATGTGTGAATATACCTTGATCTGCT	621	
Db	481	GATAATTCACACTGATGAAAAAACAAAAAGAAATGTGTGAATATACCTTGATCTGCT	540	
QY	622	TCACAGAGTACCTGAGCAAGTCAAAAAAGCCGCTTTGCTTTGATGTGAGATGTGCTT	681	

Query Match	4.88;	Score 46.2;	DB 6;	Length 1263;
Best Local Similarity	47.08;	Pred. No. 0.0068;		
Matches 178; Conservative	0;	Mismatches 198;	Indels 3;	Gaps 1.

RESULT 6
US-10-602-494-313/c
; Sequence 313, Application US/10602494
; GENERAL INFORMATION:

```

: SEQ ID NO 313
: LENGTH: 3095
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens),
US-10-602-494-113

```

Query Match	4.7%	Score 45.2;	DB 6;	Length 3095;
Best Local Similarity	50.5%;	Pred. No. 0.018;		
Matches 110; Conservative	0;	Mismatches 108;	Indels 0;	Gaps 0

QY	705	CACACAGCGCTACAGTACCAATATTTACATTGTAAAACACTAGAGAAAAAGAAAAATC	764
Db	2084	CAACCAAAAPAAACACCTCCCTCTTTTACCTCCCAAAATTCACCAAAAAACAAAAA	2022
QY	765	ATCTATATTTGATGACTACAAAGAAAAATTTAAAAACGTATCTTGATCTCAAAACAAA	824
Db	2024	AAAAAAAAAAAAAACAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	1965
QY	825	TGATTCACATTGTTCAAAGCATTATCGAAAAAGATTTCGACGACCATATCAAGT	884
Db	1964	CAAAAAAAAACAATTTTCACAAACAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTT	1905
QY	885	TAAAGACGAGCTTCCAAAATATCTTTACCCATATA	922
Db	1904	AAATCCAAACCATTCACAAACAAACAAAAAACA	1867

```

RESULT 7
US-10-257-166B-34/C
; Sequence 34, Application US/10257166B
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
; TITLE OF INVENTION: Genes Implicated in Pharmacogenomics
; FILE REFERENCE: 5013.1011
; CURRENT APPLICATION NUMBER: US/10/257,166B
; CURRENT FILING DATE: 2002-10-07
; PRIOR APPLICATION NUMBER: PCT/EP01/07470
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 34
; LENGTH: 17703
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
;
; OTHER INFORMATION: chemically treated genomic DNA
;
; US-10-257-166B-34

```

Query Match	4.63;	Score 44.6;	DB 6;	Length 17703;
Best Local Similarity	46.93;	Pred. No. 0.051;		
Matches 172; Conservative	0;	Mismatches 194;	Indels 1.	Gap 1

[illegible]

Db	15619	TAAATAATAAAAAATTAATAAATAATATTCATCATATAAAACAAAAATTAAACACAA	15660
Qy	797	AACTGTTATCTTGACTCAAAAACAAATGATTCACATTTGTTCAAGCATTTACGAA	856
Db	15559	TACCTATATTTTATTAATAAAAAACAACCTTTAATCAAAAACSTATAAAAAAAAACAAAAAA	15500
Qy	857	AGAATTTGCAAGCAGCCAAATATCAAGGTAAAGCAGCCAGCTCCAAATATCTTTACCC	916
Db	15499	ATCAATGCAATATATATTAATAAAATCAATGCACAAAAAATCCATATAAAAAATTTCTACT	15440
Qy	917	AATATAT 923	
Db	15439	AATTAAT 15433	

```

RESULT 8
US-10-603-113-3698
; Sequence 3698, Application US/10603113
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NOCLETIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/10/603,113
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: US/09/248,796
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 3698
; LENGTH: 1287
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-603-113-3698

Query Match          4.6%; Score 44.4; DB 6; Length 1287;
Best Local Similarity 48.8%; Pred. No. 0.022;
Matches 120; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

OY 381 TGAAGTGGCACTTAAAGAGGTAGCAGAACTGAAATTGCAGATGTAAGCCTTAAGAAAGC 440
DB 213 TAATTTAGAGTAATGATGAGGCTAATTAATGATTCAAAAGATGAAACATTTGGAAGAACCA 272
OY 441 CTTGATGAGTACACTCCAGATGTACCGCTCAATCATCCGCTTTAATATGAAGTAA 500
DB 273 CGGTGATGGATTCATTAAGATAGTGTATTAATGTCAGATCTTGATGACGAGATGA 332
OY 501 GGCCAAAGAAAGTCTCTGAAAAAGCCAAAGCAGAGTGCTGATTTTGCTCAATTAGCCAA 560
DB 333 AGAGCATGAGATGACGACGAGATGATGAAGACGAAAGAACAGAAAGATGAAGTACTA 392
OY 561 AGATTAATTCAACTGTATGAAAAACAAAAGAAATGCTGAGAAATTAACCTTTGATCTGC 620
DB 393 CGACATATCAGATTTAGGAACCATTAATGAAGGATTTGTTGTGAACCTTAATGAAAA 452
OY 621 TTCAC 626
DB 453 CTCAC 458

RESULT 9
US-10-603-113-5931
; Sequence 5931, Application US/10603113
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NOCLETIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/10/603,113
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: US/09/248,796
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 5931
; LENGTH: 1287
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-603-113-5931

```

```
; SEQ ID NO 5931
;
; LENGTH: 342
;
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-603-113-5931
```

	Query Match Best Local Similarity	4.6% 49.0%	Score 44.2	DB 6	Length 342
	Matches 118	Conservative	0	Mismatches 123	Indels 0
				Gaps 0	
OY	610	TTTGATTTGCTTCAACAGAGTACCTGAGCAAGTCAAAAAGCCGCTTTCGCTTAGAT	669		
Db	4	TGTCATATATGATTAATAAGGTAAATTTATTCAGAAATGAATTTCAATTTATCTAGTAAT	63		
OY	670	GTGATGGTGTCTTCTGATGTATTACACCAACTGGCACACAAGCCTACAGTAGCCAAATAT	729		
Db	64	TTAATGTGGAAATGATGTATTATCTGAGCTTTAAAAGATCATGAAATGAATTAATAT	123		
OY	730	TACATTTGTAACACCTCAGTAAGAAAAACAAAAATCATCTAATATTGATGACTACAAAGAA	789		
Db	124	AAAGAAATTTAAACCTCAAAATTAATTTATCAAAATCATTAATACAAAACCAACTCAAAAAGT	183		
OY	790	AAATTTAAAACTGTATCTTGACTCAAAAACAAATGATTTCAACATTTGTTCAAAAGCATT	849		
Db	184	GATGAAAAAAATTAATTTAATAGCTAATTAAGAAAAAAGATACTAAATGATTAATTAATTTT	243		
OY	850	A 850			
Db	244	A 244			

```

      RESULT 10
PCT-US02-38582-100/c
; Sequence 100. Application PC/TUSO238582
GENERAL INFORMATION:
APPLICANT: SAGRES DISCOVERY
APPLICANT: MORRIS, David W.
APPLICANT: ENGELHARD, Eric K.
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
TILE OF INVENTION: CANCER
FILE REFERENCE: 52945200143
CURRENT APPLICATION NUMBER: PCT/US02/38582
CURRENT FILING DATE: 2003-06-03
PRIOR APPLICATION NUMBER: 09/997,722
PRIOR FILING DATE: 2001-11-30
NUMBER OF SEQ ID NOS: 301
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 100
LENGTH: 96596
TYPE: DNA
ORGANISM: Homo Saplen
PCT-US02-38582-100

Query Match          4.5%; Score 43.8; DB 1; Length 96596;
Best Local Similarity 51.3%; Pred. No. 0.16;
Matches 102; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY       188 AACAGTCGTTGTAAATATGCACCATCCAAAAAGTTTTGGAAAACAATATGGCTCGAGGC 247
           ||| |||||
DB        7691 AAAAAAAAAATTCTTAACCAATCGTGGAIAAATAATTTGCCAAGAAGAAATGTGTATGTAT 7632Z
           ||| |||||

QY       248 TTGATGATNAAGAGGTTGATGATNACTATGCGCAAAGAAAAAACAAATATGGCCAAAACCT 307
           ||| |||||
DB        7631 CTATGTATATATTAAGTGGTTGATTAATAAAATGCAATGAAATATATATTTCTTTTGGAAATAT 757ZZ
           ||| |||||

QY       308 ACCAACGTCGTCTGTCAACAGCAGGTATGACTTGGAAACAGTAAAGCTCAAAATTCGTA 367
           ||| |||||
DB        7571 GCAAAGCAATGTAGAGAGACAAGATTAAATTCATATAGAACTTAACCTTTGTGTTCTG 7512Z
           ||| |||||

QY       368 CAAGTAATTAGTTCAGCTT 386
           |   ||| |||||
DB        7511 TATAAATATTTCTTGAGAGGT 7493

```

Other INFORMATION: chemically treated genomic DNA
US-10-257-166B-164

QY	716	ACAGTAGCAGTATTACCTTGTAAACACCCACTAGACAAACGAAAAATCATCTAAVATG	775
Pb	2050	ACAAAATCAATCTTTAACTTTAAATCTCATTTAAATTCAAAAAATATATATACCA	1991
QY	776	ATGACTACAAAGAAAAATTAAAACTGTATCTTGACTCAAAACCAAAATGATTCACAT	835
Pb	1990	TACTCCTTAAAAATCTAATAAAATTTAACTTCTTCACCTAAATTTACCGATTACAA	1931
QY	836	TTGTTCAAAGCATTTATCGGAAAGAATTGCACGACCATATTCAGTTTAAGGCCAAG	895
Pb	1930	TATTTTATACAAATTTTCAAAAAAATACACAAACCCCAATATCCAAATTAATAAACCTTA	1871
QY	896	CCTTCCAAATATCTTT	912
Pb	1870	CTATAAATCTTAAATTT	1854

```

RESULT 12
US-10-626-717-3404/C
; Sequence 3404, Application US/10626717
; GENERAL INFORMATION:
; APPLICANT: Andersen, Scott E.
; APPLICANT: Byrum, Joseph R.
; APPLICANT: De La Pena, Robert C.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(15878)D
; CURRENT APPLICATION NUMBER: US/10/626,717
; CURRENT FILING DATE: 2003-07-25
; PRIOR APPLICATION NUMBER: US 10/304,123
; PRIOR FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: US 09/594,596
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 10952
; SEQ ID NO 3404
; LENGTH: 653
; TYPE: DNA
; ORGANISM: Trifolium aestivum
; FEATURE:
; OTHER INFORMATION: Clone ID: uc-ts191316081c11b1
; US-10-626-717-3404

```

[illegible]

494-302

371 ACMAAC 1865


```

RESULT 14
US-10-640-833-1168
; Sequence 1168, Application US/10640833
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; APPLICANT: Bush, David
; APPLICANT: Zeng, Qiangong
; APPLICANT: Opperman, Timothy
; APPLICANT: Houseweat, Chad Eric
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: PATH03-15
; CURRENT APPLICATION NUMBER: US/10/640,833
; PRIORITY FILING DATE: 2003-08-14
; PRIOR APPLICATION NUMBER: US 09/583,110
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 1168
; LENGTH: 5898
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-640-833-1168

Query Match
Best Local Similarity 48.6%; Score 43; DB 6; Length 5898;
Matches 118; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 408 ACCTGAATTGACAGATGAAGCCTATTAAGAAAGCCTTGTGAGTACCTCCAGATGTAAC 467
DB 1707 AGTTGGAGATCAATAAATCAACAGACATATGACCTCTAAATAATGAAAAATTCAGAAAAAAC 1766
QY 468 GGCTCAATATCATCCGCTTAATTAATGAAGTAAGCCAAAGAGTCTCGAAAAACCCAA 527
DB 1767 TGTAGAGAAAGTCCAGTAATTAATGAAGGACAGTAGAGGTAACCTCAATCAATCAAG 1826
QY 528 GGCAGAGAGTGTGATTTGCTCAATTAAGCCAAAGATTAATCACTGATGAAAAACAAA 587
DB 1827 AACAGAAAAACCAAGTTCACAGCTGCAGAGAAACAAACAACTCTGGGAAAAATAGCTAA 1886
QY 588 AGAAAAATGTGGAGAAATTAACCTTGTGATTCGCTTCAACAGAGTACCTGAGCAATCAA 647
DB 1887 CGAAAAATCTGGAGAGATTCCAATTAACCTAGTATTCAAAAACCAAGTTGAGAAATC 1946
QY 648 AAA 650
DB 1947 AAA 1949

RESULT 15
US-10-617-320-648
; Sequence 648, Application US/10617320
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: CD/ROM ISO9660
COMPUTER: <Unknown>
OPERATING SYSTEM: <Unknown>
SOFTWARE: <Unknown>
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/617,320
FILING DATE: 10-Jul-2003
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/107,433
FILING DATE: 30-Jun-1998
APPLICATION NUMBER: 60/085131
FILING DATE: May 12, 1998
APPLICATION NUMBER: 60/051553
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Arinello, Pamela Deneke
REGISTRATION NUMBER: 40,489
TELECOMMUNICATION INFORMATION:
REFERENCE/DOCKET NUMBER: GTC-011
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 648:
SEQUENCE CHARACTERISTICS:
LENGTH: 5919 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: misc.feature
LOCATION: (B) LOCATION 1...5919
SEQUENCE DESCRIPTION: SEQ ID NO: 648:
US-10-617-320-648

Query Match
Best Local Similarity 48.6%; Score 43; DB 6; Length 5919;
Matches 118; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 408 ACCTGAATTGACAGATGAAGCCTATTAAGAAAGCCTTGTGAGTACCTCCAGATGTAAC 467
DB 1728 AGTTGGAGATCAATAAATCAACAGACATATGACCTCTAAATAATGAAAAATTCAGAAAAAAC 1787
QY 468 GGCTCAATATCATCCGCTTAATTAATGAAGTAAGCCAAAGAGTCTCGAAAAACCCAA 527
DB 1788 TGTAGAGAAAGTCCAGTAATTAATGAAGGACAGTAGAGGTAACCTCAATCAATCAAG 1847
QY 528 GGCAGAGAGTGTGATTTGCTCAATTAAGCCAAAGATTAATCACTGATGAAAAACAAA 587
DB 1848 AACAGAAAAACCAAGTTCACAGCTGCAGAGAAACAAACAACTCTGGGAAAAATAGCTAA 1907
QY 588 AGAAAAATGTGGAGAAATTAACCTTGTGATTCGCTTCAACAGAGTACCTGAGCAATCAA 647
DB 1908 CGAAAAATCTGGAGAGATTCCAATTAACCTAGTATTCAAAAACCAAGTTGAGAAATC 1967
QY 648 AAA 650
DB 1968 AAA 1970

```

Search completed: September 4, 2003, 21:56:06
Job time: 190 secs

BEST AVAILABLE COPY

THIS PAGE BLANK (USPTO)

|||||
b 241 KTKFKSSNIDYKAKLTKVLTOKRNDSTFVQSIGKELOANIKVDAQFQNFQYIG 300
y 310 GGDSSSSSSSTNE 322
|||||
b 301 GGDSSSSSSSTNE 313

ESULT 4

AMS5079

D AAMS5079 standard; Protein; 213 AA.

X AAMS5079;

X 02-OCT-1998 (first entry)

X Streptococcus pneumoniae SP0021 protein.

X Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;

X detection; pneumonia; otitis media; meningitis.

X Streptococcus pneumoniae.

X MO9818930-A2.

X 07-MAY-1998.

X 30-OCT-1997; 97WO-US19422.

X 31-OCT-1996; 96US-0029960.

X (HOMA-) HUMAN GENOME SCI INC.

X Choi GH, Hromocky J A, Johnson LS, Kunsch CA;

X WPI: 1998-272224/24.

X N-PSDB: AAV27340.

X Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae - or their epitope-containing fragments, useful in protective or therapeutic vaccines, and for diagnosis

X Claim 11; Page 55; 118pp; English.

XX The present sequence represents a protein from Streptococcus pneumoniae.
XX The nucleic acid sequence encoding the Streptococcus pneumoniae protein
XX can be useful in vaccines for inducing protective antibodies against
XX Streptococcus pneumoniae, for treatment or prevention of infection e.g.
XX pneumonia, otitis media or meningitis. Probes based on the nucleic acid
XX are used to detect Streptococcus infection (by usual hybridisation or
XX amplification methods), also for isolating Streptococcus genes or their
XX allelic variants. The protein can be used similarly for detecting or
XX antibodies in standard immunoassays, especially for diagnosing or
XX monitoring infections. Antibodies which bind the protein are used to
XX detect corresponding antigens, to purify the protein and for passive
XX immunisation (optionally coupled to a toxin). Vaccines are administered,
XX e.g. by injection, orally or through the skin, typically at 0.01-1000
XX (especially 10-300) mu g/ml per dose.

X Sequence 213 AA;

X Query Match 57.7%; Score 916; DB 19; Length 213;

X Best Local Similarity 93.6%; Pred. No. 3.3e-62;

X Matches 190; Conservative 0; Mismatches 5; Indels 8; Gaps 2;

31 SKGSEGAADLSMKGVITEHQFYEQVKNPSAQOVLNNTIKVFEKQYSGSELDKREVD 90
|||||
1 SKGSEGAADLSMKGVITEHQFYEQVKNPSAQOVLNNTIKVFEKQYSGSELDKREVD 60
|||||
91 TIAEEKQYGENQVRLSQAQMTLETRKQIRTSKLVKVAEELTDEAYKKADE 150
|||||
61 TIAEEKQYGENQVRLSQAQMTLETRKQIRTSKLVKVAEELTDEAYKKADE 120

Fri Sep 5 09:24:50 2003

us-10-049-4

QY 151 YTPDVTAAQIIRLNNEDEKAKEVLEKAKAEGADFAQLAKNDSTDEKRENGCEITPDSASTE 210
|||||
Db 121 YTPDVTAAQIIRLNNEDEKAKEVLEKAKAEGADFAQLAKNDSTDEKRENGCEITPDSASTE 180
QY 211 VP-EQVKAFA-----LDVD 225
|||
Db 181 VPGASPKPLFAFRGCMFLDVD 203

RESULT 5

ID ABP54573 standard; Protein; 213 AA.

X ABP54573;

X 04-SEP-2002 (first entry)

X S. pneumoniae SP021 protein sequence SEQ ID NO:34.

X Streptococcus pneumoniae; epitope; vaccine; antigenic protein;

X antibacterial; Streptococcal infection; detection.

X Streptococcus pneumoniae.

X US2002061545-A1.

X 03-MAY-2002.

X 22-JAN-2001; 2001US-0765272.

X 30-OCT-1997; 97US-0961083.

XX (CHOI/) CHOI G H.
XX (KUNS/) KUNSCH C A.
XX (BARA/) BARASH S C.
XX (DILL/) DILLON P J.
XX (DOUG/) DOUGHERTY B.
XX (FANN/) FANNON M R.
XX (ROSE/) ROSEN C A.

XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX PI
XX Rosen CA;

XX WPI: 2002-479261/51.
XX N-PSDB: AB084808.

XX New Streptococcus pneumoniae antigens, useful for detecting
XX Streptococcus and for preventing or attenuating disease caused by
XX Streptococcus infection -
XX
XX Claim 11; Page 24; 70pp; English.

XX AB084792 to AB084904 represents nucleic acids which encode the
XX Streptococcus pneumoniae antigens given in ABP54557 to ABP54659.
XX The S. pneumoniae antigens have antibacterial activity and can be
XX used in vaccines. The S. pneumoniae antigens can also be used to
XX prevent or attenuate a Streptococcal infection in an animal. The
XX polynucleotides encoding the S. pneumoniae antigens can be used to
XX detect Streptococcus nucleic acids. AB084905 to AB085130 represent
XX primers used in the cloning of S. pneumoniae ORFs (open reading frames)
XX which are used in an example from the present invention.

X Sequence 213 AA;

X Query Match 57.7%; Score 916; DB 23; Length 213;

X Best Local Similarity 93.6%; Pred. No. 3.3e-62;

X Matches 190; Conservative 0; Mismatches 5; Indels 8; Gaps 2;

31 SKGSEGAADLSMKGVITEHQFYEQVKNPSAQOVLNNTIKVFEKQYSGSELDKREVD 90
|||||
1 SKGSEGAADLSMKGVITEHQFYEQVKNPSAQOVLNNTIKVFEKQYSGSELDKREVD 60
|||||
91 TIAEEKQYGENQVRLSQAQMTLETRKQIRTSKLVKVAEELTDEAYKKADE 150
|||||
61 TIAEEKQYGENQVRLSQAQMTLETRKQIRTSKLVKVAEELTDEAYKKADE 120

BEST AVAILABLE COPY

THIS PAGE BLANK (USPTO)

FILE 8
8-858-207A-508
Sequence 508, Application US/08858207A

GENERAL INFORMATION: Black, Michael
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Nicholas, Richard
APPLICANT: Stoddola, Robert
TITLE OF INVENTION: No. 6348328el Compounds
NUMBER OF SEQUENCES: 352
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Smithkline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: Fastiso for Windows Version 2.0
CURRENT APPLICATION DATA: US-08/858, 207A

APPLICATION NUMBER: 22-00000000
FILING DATE: 09-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIORITY NUMBER: 60/017670

APPLICATION NUMBER: 60/017670
FILING DATE: 14-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimm, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50475
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090

TELE: INFORMATION FOR SEQ ID NO: 508
SEQUENCE CHARACTERISTICS:
LENGTH: 74 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: NO. 634832be
PS-08-856-207A-508

	20 38;	Score 322; DB 4; Length 74:
Query Match	97.18;	Pred. No. 1,2e-19;
Best Local Similarity		
Matches	66; Conservative	1; Mismatches 1; Indels 0; Gaps 0.
	10 MKKKLLGATITLTVALLAACSGKSEADILISKRGVITEHOFYEQVCNPSAQOYLLNM	69
Oy		
	1 MKKKLLGATITLTVALLAACSKGSEADILISKRGVITEHOFYEQVCNNPSSAOXVLNM	60
Db		

RESULT 9
US-09-134-001C-3811
Sequence 3811, Application US/09134001C

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: SIC 00,
CURRENT APPLICATION NUMBER: US/p9/134,001C

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR APPLICATION NUMBER: US
FILING DATE: 1997-08-14

NUMBER OF SEQ
; CEO TD NO 381

LENGTH: 330
TYPE: PRT

US-09-134-001C-3811

Query Match

Matches 9

23

5

2 5 7

1300v

Db 13

QY 17

Db 18

23 QY

DB 24

2

RESULT 10
MS-09-252-99

; Patent No

APPLICANT ;
STATE OF

TITLE OF
ATTN: REF

CURRENT
CURRENT

PRIOR FI

PRIOR FI
NUMBER C

; SEQ ID NO

THIS PAGE BLANK (USPTO)

BEST AVAILABLE COPY

THIS PAGE BLANK (USPTO)